Comparing Measures of Average Color Fidelity

Michael P Royer[1]

[1]Pacific Northwest National Laboratory
620 SW 5th Avenue, Suite 810
Portland, OR 97204
michael.royer@pnnl.gov

This is an archival copy of an article published in *LEUKOS.* Please cite as:

Royer MP. 2017. Comparing Measures of Average Color Fidelity. LEUKOS. 14(2):69-85. DOI: 10.1080/15502724.2017.1389283.

PNNL-SA-128752

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.
(8/2010)

Abstract

The introduction of new measures of color rendition, especially IES TM-30-15, has stirred debate within the lighting industry on the relative merits of the tools, as well as the amount of difference between the new tools and prior tools, such as CIE R_a. This article focuses on comparing three measures of average color fidelity: IES R_f, CIE R_a, and CIE R_f. Using a large set of commercially-available, experimental, and theoretical spectral power distributions (SPDs), the analysis contrasts past efforts to make similar comparisons using smaller or more focused datasets. It highlights the interactive effect of gamut shape and color space non-uniformity, which results in a range of IES R_f values of at least to 50 to 86 for SPDs having a CIE R_a value of 80. It also examines how these differences can be overlooked in psychophysical experiments relying on a small number of SPDs, which can present misleading findings on the value and meaning of the measures. When considering the results, it is important to remember that average color fidelity is only one aspect of color rendition.

1 Introduction

The proposal of numerous measures of color rendition over the past decade has resulted in multiple attempts to compare them. These comparisons have been made in two main ways: (1) numerical analyses focused on calculated values for different measures of color rendition using a set of spectral power distributions (SPDs) [David and others 2015; Davis and Ohno 2010; Houser and others 2013; Khanh and others 2016c; Rea and Freyssinier 2015; Royer and others 2017a; Royer and Wei 2017; Wei and Houser 2012], and (2) examining correlations between one or more measures of color rendition and data from human evaluations of objects illuminated by a small set of lighting conditions [Islam and others 2013; Jost-Boissard and others 2014; Khanh and Bodrogi 2016; Khanh and others 2016a; 2016b; Royer and others 2016; Smet and others 2011; Xu and others 2016]. This article examines the methods used in past studies, and presents new numerical comparisons of selected measures of color rendition. The results counter some previously drawn conclusions, and provide insight for understanding measures of color rendition, designing psychophysical experiments, engineering new products, and writing performance specifications.

This article focuses on comparing the IES TM-30-15 [IES 2015] Fidelity Index (IES R_f) to the CIE General Color Rendering Index (R_a) [CIE 1995]. Additional analysis is provided to compare IES R_f (from IES TM-30-15) to CIE R_f, which was adopted in CIE 224:2017 [CIE 2017]. These three measures were selected because they are formalized by a national or international lighting authority; other measures of color fidelity that have been proposed (for example CQS Q_f [Davis and Ohno 2010], CRI2012 [Smet and others 2013], and $R_{a1\text{-}14}$ [Khanh and others 2016c]) are not included because they are neither adopted nor being actively considered by a lighting authority. In addition to examining the magnitude of the differences between the measures, the specific causes of the differences are examined. Most importantly, gamut shape [Royer and others 2017a] is shown to be an important correlate of difference between CIE R_a and IES/CIE R_f, highlighting the need to consider sources of varying gamut shape when investigating color rendition.

While this article focuses on measures of average color fidelity, complete specification and understanding of color rendition requires a more comprehensive set of tools. Chroma shift and gamut shape have been covered by Royer, Houser, and David [2017]. Gamut area is planned to be covered in a separate article.

1.1 Past Comparisons of IES R_f and CIE R_a

IES R_f and CIE R_a were first compared in published literature by David and colleagues [2015]. Using a set of 401 SPDs from Houser and colleagues [2013], the relationship between IES R_f and CIE R_a was demonstrated, but statistics such as r^2 were eschewed because, as noted, relying on correlation alone can mask large differences that have a substantial practical effect. Instead, the analyses by David and colleagues focused on examining the potential range of differences in one measure at any given value of the other measure. For example, at a CIE R_a value of approximately 80, the 401 SPDs examined had a range in IES R_f values of about 71 to 86.

As David and colleagues noted, the set of SPDs used was not a representative sample; a majority were real light sources of various types, with the remainder mostly being models of currently-realizable sources. Establishing a set of sample SPDs is a challenge for anyone trying to numerically compare measures of color rendition, as there is no feasible way to derive a set that is representative of the lighting market—either installed or on the market. Even if such a set of existing light sources could be compiled, it would not be maximally informative, because it does not include potential future light sources. Perhaps the most important statistic is how large of a possible "error" could be made by an inaccurate measure, since color rendition measures are also used to aid in the development of future light sources.

Not long after IES TM-30-15 was published, some began to suggest that IES R_f and CIE R_a were not very different, and thus there was no reason to adopt a new measure. One such argument was made by Rea and Freyssinier [2015], who reported the coefficient of determination (r^2) between IES R_f and CIE R_a for "commercially available broad-band, white light sources" was 0.95. Another article simply stated, without supporting evidence, that R_f was not different enough to merit adoption, as it may "create upheaval" in the lighting industry [Teunissen and others 2016]. Khanh, Vin, and Bodrogi [2016c] suggested that "fidelity metrics generally correlate well among each other," with differences based on correlated color temperature (CCT). These statements may have influenced public opinion of IES R_f; however, this paper contends that the underlying assumptions and analyses for these statements are incomplete or inaccurate. Some of these statements also ignore major components of IES TM-30-15 that cover aspects of color rendition beyond average color fidelity, but that issue is not the focus of this article.

David and colleagues [2015] identified that differences between CIE R_a and IES R_f can be related to the presence of sharp changes in the SPDs; that is, broadband SPDs, like incandescent or phosphor-coated LEDs, tend to exhibit less difference between CIE R_a and IES R_f than highly-structured SPDs, such as tri-phosphor fluorescent or typical color-mixed LEDs. This is due to the difference in wavelength sensitivity of the color sample sets [David and others 2015; Smet and others 2015]. Because CIE R_a is calculated using only eight Munsell color samples, there are particular wavelength regions with increased sensitivity and others with little or no sensitivity. Small changes in sensitive regions, created with highly-structured SPDs but not broadband SPDs, can have large influences on CIE R_a values. The same is not true of IES R_f, because the color evaluation samples have near-uniform wavelength sensitivity in aggregate. This demonstrates that an argument such as that of Rea and Freyssinier is incomplete, because the analysis only included commercially-available, broadband light sources.

Smet and colleagues [2015] effectively demonstrated specific causes of disparities between CIE R_a and IES R_f, focusing on the number of samples, color space uniformity, and sample set wavelength uniformity. Their analysis relied on a set of 139 SPDs for commercially-available LED and fluorescent products. Combined, average differences between the two measures for broadband and narrowband light sources were about 2 and 5 points, respectively, with individual differences as high as 14 points. Average disparities arising from the different color spaces were about 1 and 3 points (with a maximum of about 6 points) for the same two groups. Disparities from the different sample set uniformity were on average 1 and 2 points (up to 6 points). Many of the differences alone are relatively small, but may still be practically important due to the extensive use of thresholds for establishing color rendition criteria, and cast doubt on arguments that CIE R_a and IES R_f are not meaningfully different. Importantly, Smet and colleagues only examined a limited set of commercially-available SPDs.

As the literature on color rendition has accumulated, it has become clear that simple average characterizations of color rendition are insufficient for addressing human subjective evaluations of color rendition [de Beer and others 2015; Esposito 2016; Ohno and others 2015; Royer and others 2017a; Royer and others 2017b; Royer and others 2016; van der Burgt and van Kemenade 2010; Wei and others 2016a; 2016b]. In particular, gamut shape [Royer and others 2017a] has been identified as a critical concept for predicting color preference and other qualities. Until the publication of IES TM-30-15, however, there was no method for characterizing gamut shape recognized by a standard-writing lighting organization. As a result, and in combination with energy efficiency goals, common commercially-available light sources offer little variation in gamut shape, tending to increase chroma for yellow-green and decrease chroma for reds [Royer and others 2017a]. Thus, as shown in this analysis, datasets relying exclusively or predominantly on commercially-available light sources have strong limitations for comparing measures of color rendition, especially when finding the extent of potential differences is important.

2 New Comparison of Average Fidelity Measures

2.1 Large SPD Set

To get a more complete picture of the differences between CIE R_a and IES R_f, a large set of 4,945 SPDs with varied gamut shape was compiled. It includes:

- 211 commercially-available light sources from the TM-30-15 Calculator Tool Library, including:
  - 17 Fluorescent Broadband
  - 22 Fluorescent Narrowband
  - 20 High-Intensity Discharge
  - 14 Incandescent/Filament
  - 128 LED Phosphor
  - 1 LED Color-Mixed
  - 8 LED Hybrid
  - 1 Plasma
- 152 experimental LED SPDs. These light sources were used in experiments at the National Institute of Standards and Technology (NIST) and Pacific Northwest National Laboratory (PNNL). The NIST sources, created with a custom-build multi-channel LED apparatus, are included in the IES TM-30-15 library. The PNNL sources were from two experiments [Royer and others 2017b; Royer and others 2016], and were created with a group of ETC Source 4 Series 2 Lustr luminaires.
- 4,582 randomly-generated theoretical SPDs, derived during the IES TM-30-15 development process, each comprised of up to four Gaussian components with varying full-width-half maximum values (as small as 1 nm). All of the SPDs have a chromaticity on the Planckian locus, with CCT between 2500 K and 6500 K.

2.2 IES R_F versus CIE R_a

Figure 1 shows the relationship between IES R_f and CIE R_a for the large SPD set, displaying only values above 50 for clarity. (Differences below 50 may also be influenced by the transformation that is applied to R_f to avoid negative values). The large SPD set covers a wide range of possible color rendition conditions, resulting in a great extent of possible differences between the two measures of average color fidelity. At a CIE R_a value of 80, the range in IES R_f values is approximately 50 to 86, with an average of 75.4. This occurs despite a moderately strong correlation between the two measures (for all 4,945 SPDs) of $r^2 = 0.78$. The coefficient of determination (r^2) for SPDs with IES R_f and CIE $R_a \geq 50$ (as shown in Figure 1) is 0.61. To reiterate, correlation is not a meaningful statistic when comparing measures of color rendition capturing the same quality (for example, average color fidelity), as previously noted by David and colleagues [2015]. Because the theoretical SPDs in this dataset only include SPDs with chromaticity on the Planckian locus, the maximum possible difference at any given value of CIE R_a is probably even greater.

It is also apparent from **Figure 1** that the set of 211 commercially-available light sources—like similar sets used in other analyses—is not representative of all possible outcomes in comparing measures of average color fidelity. While comparisons using only commercially-available light sources may be relevant at present, they provide no indication of how the measures will perform in the future, when

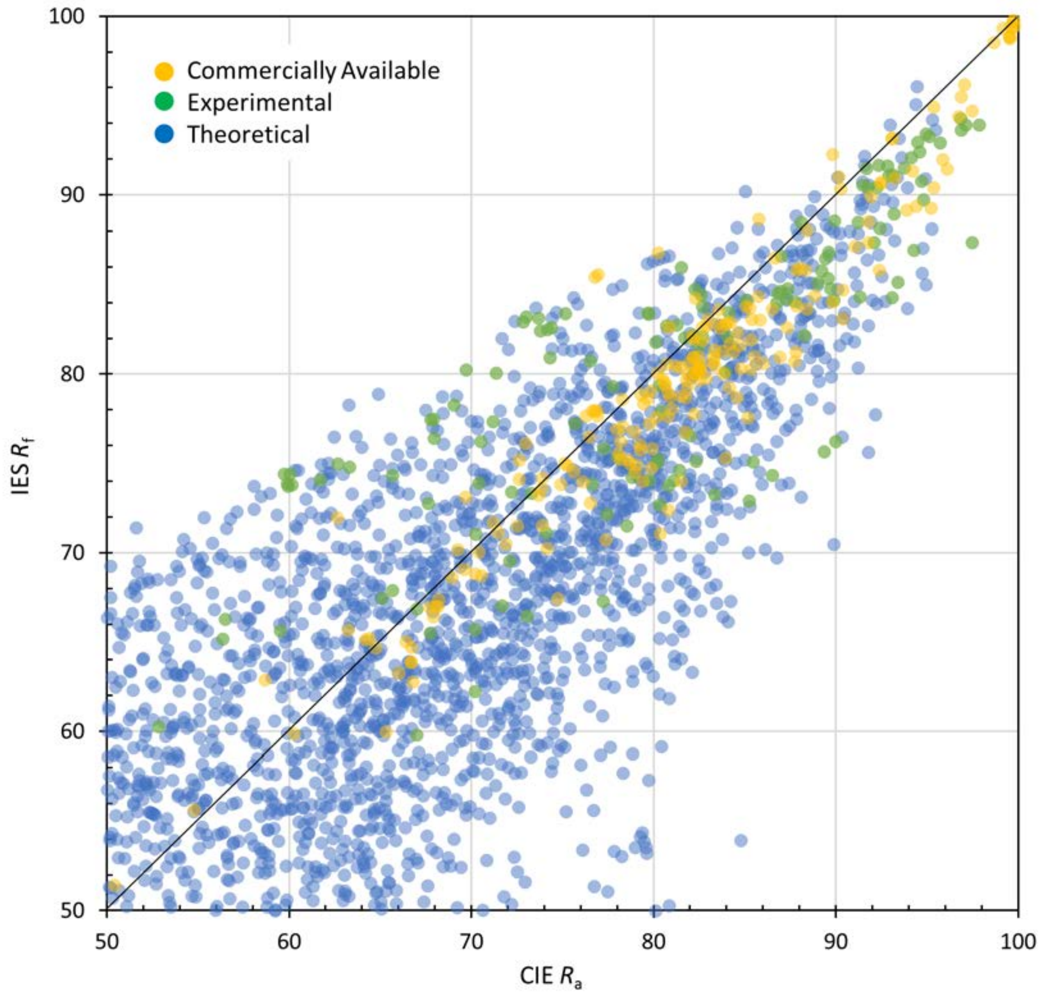

Figure 1. Comparison of CIE R_f and CIE R_a for the large SPD set, showing only data points where both measures are greater than or equal to 50. This includes 2,303 SPDs (203 commercial, 136 experimental, 1,964 theoretical).

light sources are likely to have more variety in gamut shape as spectral engineering becomes more practical. If a goal of color rendition measures is to enable spectral optimization during the engineering of new products, it is important that the measures apply to a wide range of light sources, and not just what is available today.

2.3 CIE R_f versus IES R_f

Recently, CIE 224-2017 documented a color fidelity metric based on IES R_f, keeping the designation R_f [CIE 2017]. There were three changes made to the calculation method:

1. The extrapolation method for the subset of the 99 color evaluation samples (CES) that originally did not include data outside of 400 to 700 nm was changed from a custom logarithm-based extrapolation method [David and others 2015] to a flat extrapolation. Because color matching functions are near-zero below 400 nm and above 700 nm, this has no practical effect on the subsequently calculated color rendition measures. For the large

SPD set described above, the randomly distributed range in value difference from this change alone is -0.02 to 0.004 points, with a mean of 0.00 points (**Figure 2a**).

2. The blending range for the reference illuminant was changed from 4500-5500 K to 4000-5000 K, in order to preserve the inclusion of CIE D_{50} as a reference illuminant. Again, this has minimal effect on values—and only for light sources between 4000 K and 5500 K. For the large SPD set, the randomly distributed range in value difference from this change alone is -1.80 to 1.64 points, with a mean of -0.02 points (**Figure 2b**).
3. The scaling factor applied to the average color difference was changed from 7.54 to 6.73. For IES TM-30-15, the scaling factor was determined by equalizing the average R_f and R_a values for the CIE F Series illuminants, relying on historical precedence from the calculation of CIE R_a and CQS. For CIE R_f, the scaling factor was chosen to equalize the average CIE R_a and CIE R_f values based on the commercially-available light sources in the IES TM-30-15 calculation tool library, excluding incandescent sources. The change increases all values, but does not change the rank order of the light sources. The real impact of this change relates to establishing and meeting thresholds that can be used to qualify products. A product that has an IES TM-30-15 R_f value of 80 would have a CIE R_f value of 82.1 (**Figure 2c**).

Figure 2d shows the composite effect of the three changes, whereas **Figure 3** shows CIE R_f versus CIE R_a. The change in scaling factor for CIE R_f results in more normally distributed differences centered on no difference versus CIE R_a for high-color-fidelity SPDs, which can be observed by comparing **Figures 1 and 3**. For the entire large SPD set, CIE R_f has a mean difference versus CIE R_a of 10.2 points (with CIE R_f greater than CIE R_a, on average), whereas the difference between IES R_f and CIE R_a and is only 5.36 points (with IES R_f greater than CIE R_a, on average). This is due to smaller differences among low-color-fidelity SPDs when IES R_f is used in lieu of CIE R_a. These differences among low-color-fidelity SPDs are influenced by the transformation applied to IES or CIE R_f values.

As shown in Figures 1 and 3, as average fidelity increases, the mean difference between R_f and R_a decreases. When considering only SPDs with CIE $R_f \geq 50$ and CIE $R_a \geq 50$ (as shown in Figure 3), the mean difference between CIE R_f and CIE R_a is 0.57 points, compared to -2.15 points for IES R_f versus CIE R_a (when only SPDs with IES $R_f \geq 50$ and CIE $R_a \geq 50$ are considered, as shown in Figure 1). When considering only SPDs with CIE $R_a \geq 80$, the mean difference between CIE R_f and CIE R_a is -2.7 points, compared to -4.7 points for IES R_f versus CIE R_a, with CIE R_a being less than R_f in both cases. In practical terms, the CIE R_f applies a slight "grade inflation" to IES R_f, which likely only has an effect on threshold criteria. The subsequent analysis in this report is based on IES R_f, but would not be substantially different if CIE R_f were used instead.

At press, the IES color committee is considering updating IES TM-30 to unify IES R_f and CIE R_f. Note that changing the scaling factor in R_f would not affect the non-fidelity measures included in IES TM-30, such as the Gamut Index (R_g) or Local Chroma Shift ($R_{cs,hj}$) [Royer and others 2017a]. It will affect the Local Fidelity values ($R_{f,hj}$). The changes to the CES extrapolation method and reference blending region will have a negligible effect on all IES TM-30 values. IES R_g, and the relationship between fidelity and gamut values, are discussed in Part 2 of this work.

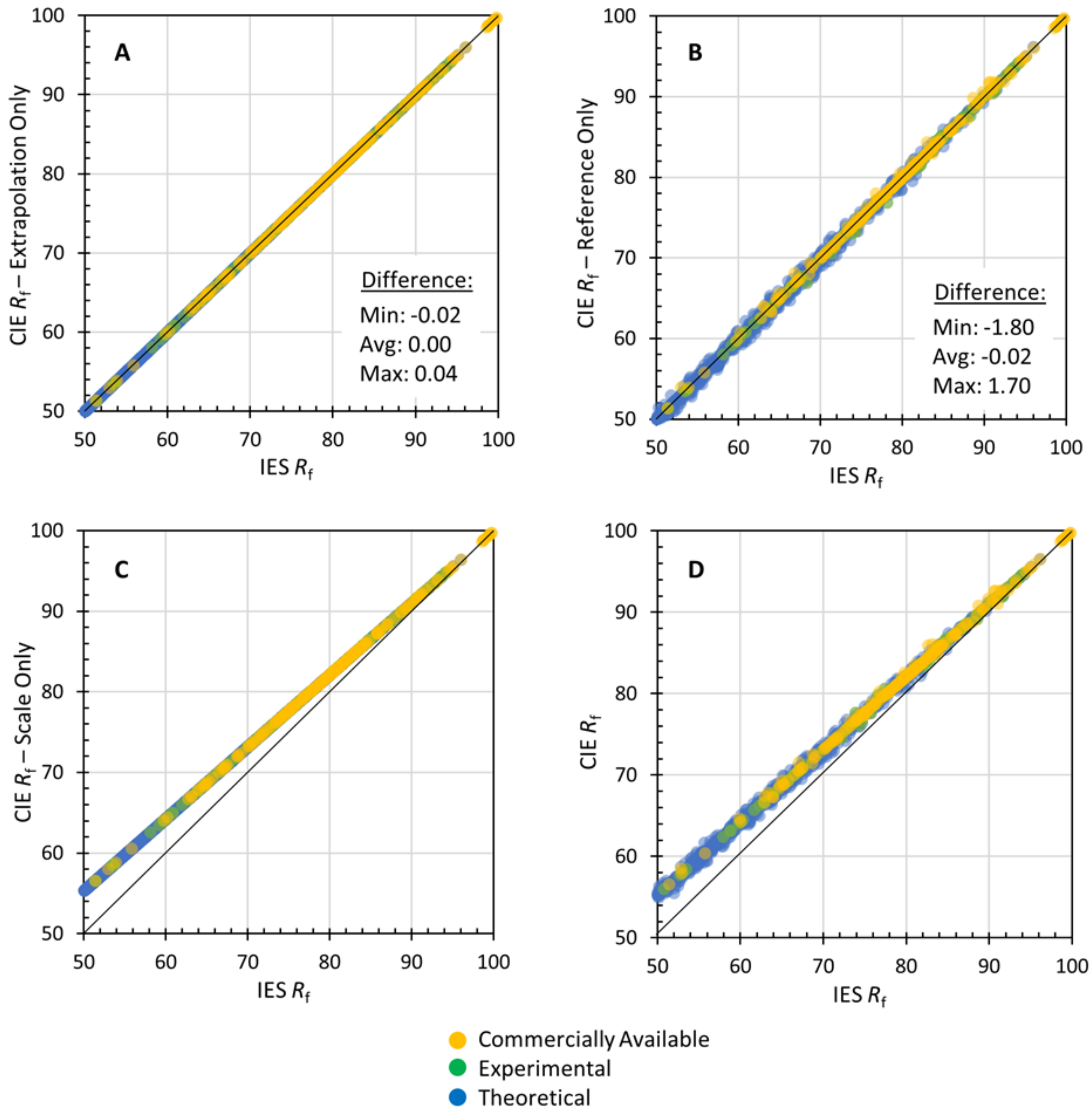

Figure 2. Comparison of CIE R_f to IES R_f. A: Result of only the change in color sample extrapolation method. B: Result of only the change in reference (applicable only to SPDs with CCTs between 4001 K and 5499 K). C: Result of only changing the scaling factor. D: Composite effect of three changes.

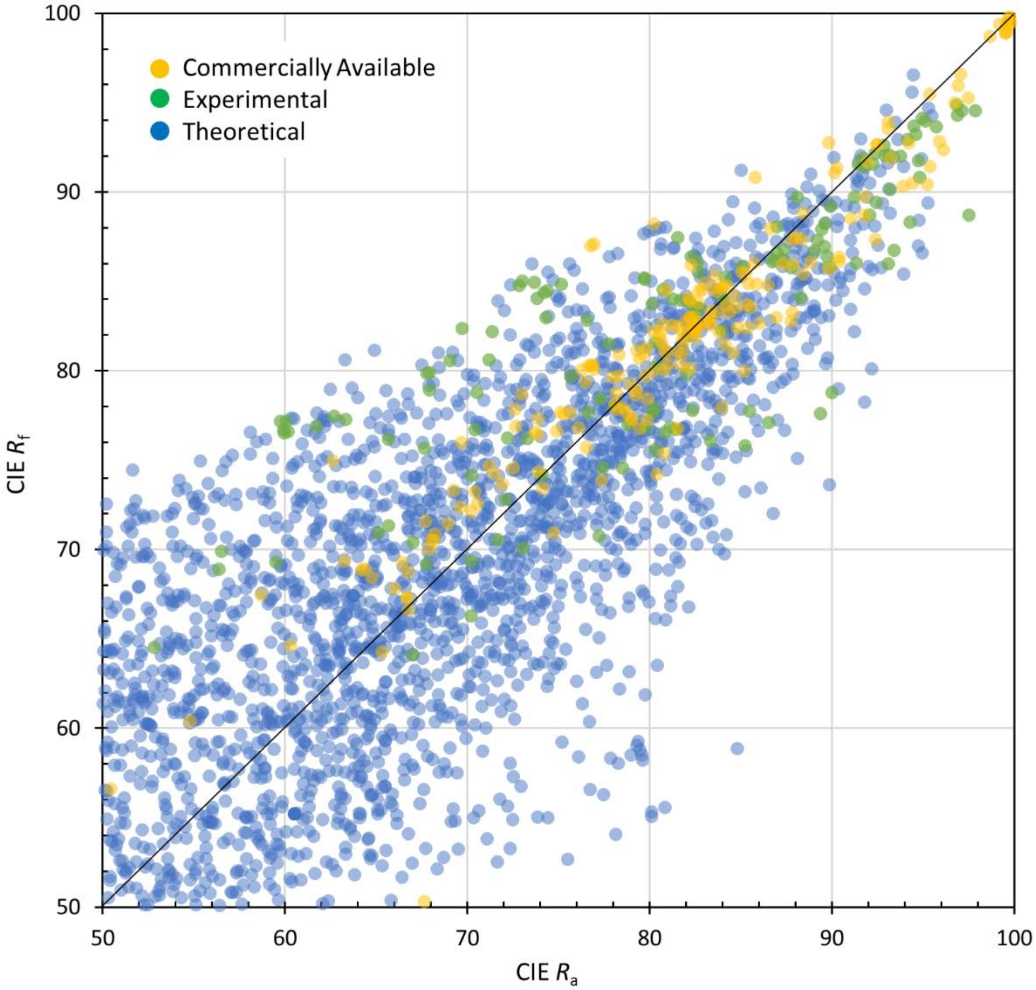

Figure 3. Comparison of CIE R_f and CIE R_a for the subset large SPD set with both measures having values greater than or equal to 50. This includes 2,495 SPDs (204 commercial, 136 experimental, 2,155 theoretical).

3 Discussion

3.1 The Role of Gamut Shape

The large SPD set can also reveal a very important difference between IES R_f and CIE R_a: CIE R_a values tend to be lower than IES R_F values for light sources that increase red chroma. Of the 2,965 SPDs in the large SPD set with $R_{cs,h1} \geq 0\%$, 74.9% had IES R_f values greater than the corresponding CIE R_a values. In contrast, of the 1,980 SPDs in the large SPD set with $R_{cs,h1} < 0\%$, only 17.7% had IES R_f values greater than the corresponding CIE R_a values. Likewise, of the 3,253 SPDs in the large SPD set with $R_{cs,h2} \geq 0\%$, 71.7% had IES R_f values greater than the corresponding CIE R_a values, compared to 14.2% of the 1,692 SPDs in the large SPD set with $R_{cs,h2} < 0\%$. Given that a vast majority of commercially-available light sources have gamut shapes that decrease the chroma of red objects versus the reference [Royer and others 2017a], these datasets provide only a partial picture when comparing measures of color rendition.

Relationships between Local Chroma Shift values ($R_{cs,h2}$, $R_{cs,h5}$, $R_{cs,h9}$, and $R_{cs,h13}$) and the difference between IES R_f and CIE R_a are illustrated in **Figure 4**. Local Chroma Shift in hue-angle bin 2 ($R_{cs,h2}$) was chosen over other Local Chroma Shift values for nominally red hue-angle bins because it displayed the strongest correlation. Note that CIE Test Color Sample (TCS) 1 would be in IES TM-30-15 hue-angle bin 2 for typical SPDs. (TCS 9 is typically near the border between hue-angle bins 1 and 2.)

For SPDs that increase red chroma (in this case, $R_{cs,h2} \geq 0\%$), there is a strong positive correlation between $R_{cs,h2}$ and the difference between IES R_f and CIE R_a ($r^2 = 0.83$). There is a weak negative correlation for the same relationship when $R_{cs,h2}$ is less than zero. This pattern does not exist for $R_{cs,h5}$ (nominally yellow chroma shift) or $R_{cs,h13}$ (nominally blue chroma shift). A similar, but weaker, pattern occurs with $R_{cs,h9}$, ostensibly because increases in red chroma and green chroma versus the reference illuminant tend to occur simultaneously. This illustrates a fundamental, systematic difference between IES R_f and CIE R_a, where SPDs with particular features are rated differently. That is, the differences between IES R_f (or CIE R_f) and CIE R_a are not random, and the use of one measure versus the other is likely to drive product development in a particular direction, as discussed in the next section.

This average relationship (linear correlation) between $R_{cs,h2}$ and the difference between IES R_f and CIE R_a values is a function of the different color spaces used for the two measures: CIE 1964 U*V*W for CIE R_a and CAM02-UCS for IES TM-30-15. CIE 1964 U*V*W is known to not accurately represent the human visual system (that is, it is non-uniform), and has been superseded by other systems. This limitation of CIE 1964 U*V*W* may be familiar to end-users in the scale of meaningfulness for CIE R_9 values.

To further investigate and separate the roles of color space uniformity and sample set wavelength uniformity on the relationships between IES R_f and CIE R_a, a third measure was calculated using the 8 TCS from CIE R_a within the exact calculation framework of IES R_f (that is, using CAM02-UCS and the IES TM-30-15 reference scheme). This measure is subsequently denoted R_f8. As opposed to the average trend, the variation in differences between IES R_f and CIE R_a for any given value of $R_{cs,h2}$ (that is, the vertical spread seen in Figure 4) includes at least three other contributing factors:

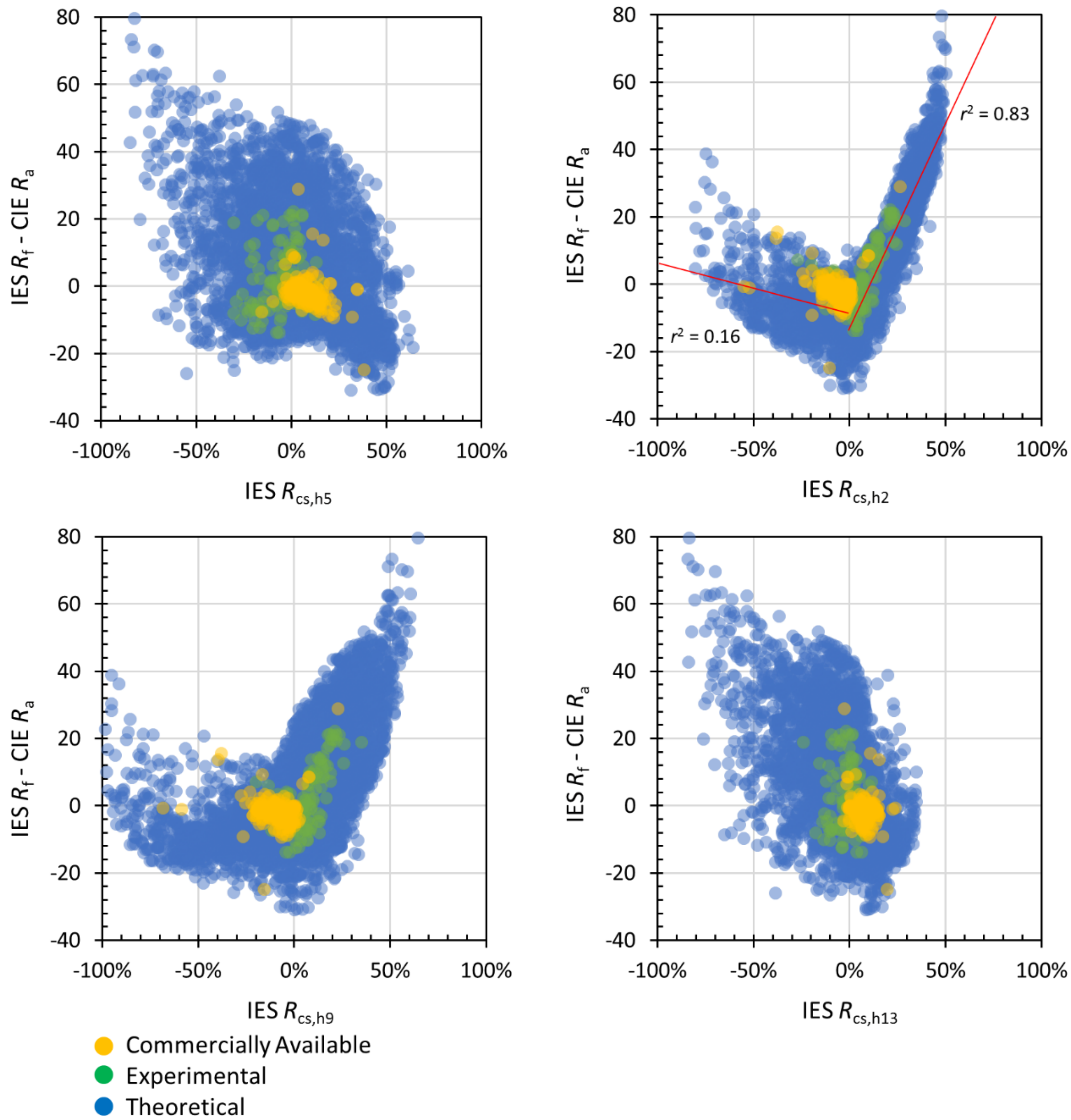

Figure 4. Difference between IES R_f and CIE R_a values versus Local Chroma Shift values for hue angle bins 2 (nominally red), 5 (nominally yellow), 9 (nominally green), and 13 (nominally blue). The strongest relationship occurs for $R_{cs,h2}$.

1. Local Chroma Shift values ($R_{cs,hj}$) are a limited characterization of gamut shape, and all hues contribute to average fidelity values.
2. CIE R_a values are derived from eight test color samples (TCS), which are not evenly distributed in the color volume or in the hue-chroma plane. The shift of any one sample (for example, TCS1) compared to the reference can be substantially different from with the average shift of samples in the same hue-angle bin, as shown in **Figure 5** for both CIE U*V*W* (left) and CAM02-UCS (right).
3. Similarly, individual SPDs interact differently with the disparate wavelength uniformity profiles of the 99 CES of IES TM-30-15 and the 8 TCS of CIE R_a [David and others 2015; Smet and others 2015], which are shown in **Figure 6**.

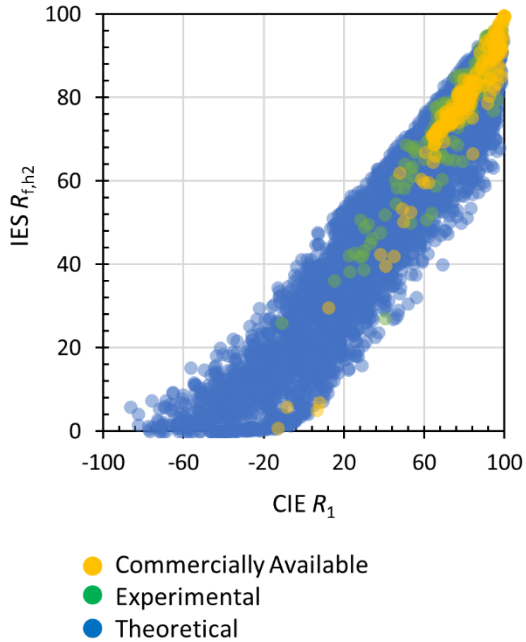

Figure 6. Local Color Fidelity for hue-angle bin 2 ($R_{f,h2}$), representing the average shift within the hue-angle bin, versus individual sample color fidelity (CIE R_1, R_f8 R_1), for the sample in the CIE Test Color Method that falls in hue angle bin 2 (TCS1). R_f8 is a measure combining the calculation framework of IES TM-30-15 (that is, using CAM02-UCS) with the color samples of CIE R_a.

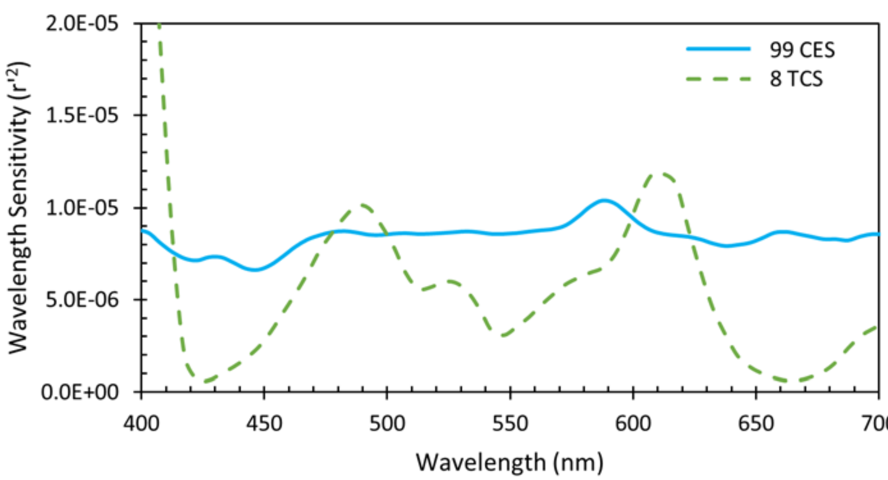

Figure 5. Wavelength sensitivity (r'^{2}) of the 99 IES TM-30-15 CES versus the 8 TCS used to calculate CIE R_a. Wavelength sensitivity describes the sample sets' sensitivity to changes in SPD.

Using the large SPD set, R_f8 is compared to both IES R_f and CIE R_a in **Figure 7**. Smet and colleagues noted that sample set wavelength uniformity contributed up to 6 points in difference between IES R_f and CIE R_a values. **Figure 7a**, comparing measures with equal calculation methods but different sample sets (R_f8 versus IES R_f, shows that the difference between measures can be larger when experimental or theoretical SPDs are also considered than when only commercial sources are

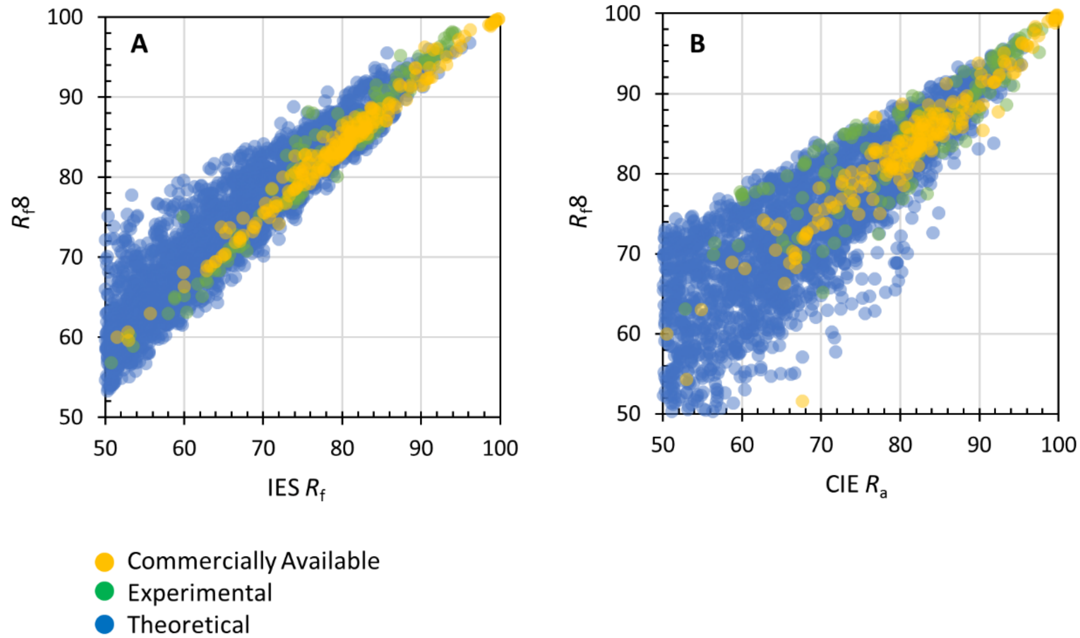

Figure 7. R_f8 versus IES R_f (A) and CIE R_a (B). The two comparisons show relative effects of sample set/wavelength uniformity (A) and color space uniformity (B).

considered. At an IES R_f value of 80, the range in R_f8 is approximately 80 to 90 with greater variation as IES R_f is reduced. Note also that R_f8 values are always higher than R_f values, demonstrating that, with this dataset, it is not possible to have a higher average color shift for the eight TCS than for the 99 CES. Over the large SPD set, this difference is generally smaller in magnitude than the difference resulting from the change in calculation methodology (**Figure 7b**)—essentially the color space and chromatic adaptation components. At a CIE R_a value of 80, R_f8 varies from approximately 65 to 89.

The absolute range in differences shown in **Figure 7a** is a function of the overall sensitivity of the sample sets. Accordingly, it can be seen that the need to adjust scaling factors between measures of color rendition is at least in part a result of the differences in sample set properties, with some also due to differences in the color spaces. One important property is the chroma of the samples, because higher chroma samples tend to have lower color fidelity when evaluated across large sets of SPDs. As shown in **Figure 8**, which compares average fidelity values for the 99 CES binned by chroma level across the large SPD set, the relationship is not linear. Although the mean chroma of the 8 TCS and 99 CES is approximately the same (under a 4000 K Planckian radiator), at 30.7 (R_f8) and 31.0 (IES R_f), all R_f8 values are higher than or equal to the respective IES R_f values, because of the specific samples included and the aggregate wavelength uniformity. The relationship between R_f8 and IES R_f showed no obvious correlation with gamut shape, as expressed via Local Chroma Shift values. As expected, differences in values for R_f8 and CIE R_a were correlated with gamut shape/Local Chroma Shift values in the same manner as the differences in values for IES R_f and CIE R_a.

The difference between IES R_f and CIE R_a values and its relationship to the difference between IES R_f and R_f8 is further illustrated by comparing standard CVGs to ones generated from the R_f8 calculation (**Figure 9**). Simply, the 8 TCS used to calculate CIE R_a do not always predict the same hue-specific

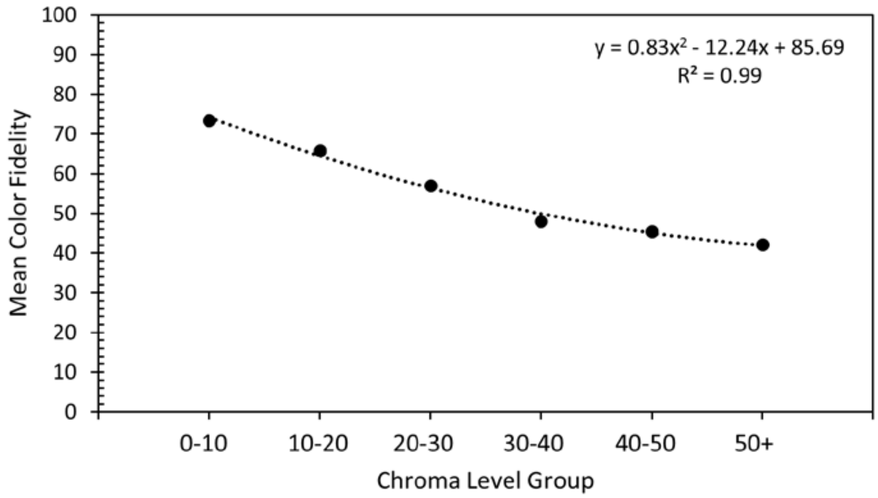

Figure 8. Mean color fidelity over the large SPD set for six groups of the 99 CES divided based on their chroma under a 4000 K Planckian radiator.

color shifts as the 99 CES used to calculate IES R_f. The difference in values between IES R_F and CIE R_a are subsequently related to how sample set differences lead to different gamut shapes, which are valued differently based on the color space, as previously described. That is, all the factors in the differences between IES R_f and CIE R_a values are cumulative and interrelated.

Smet and colleagues previously demonstrated that the non-uniformities of the CIE U*V*W* color space are CCT dependent [Smet and others 2015]. Whereas errors in both axes of hue-chroma plane in CAM02-UCS are consistent, *a*' errors in CIE U*V*W* decrease with increasing CCT, whereas *b*' errors increase. The net errors acriss all three dimensions of color space are fairly similar between CAM02-UCS and CIE U*V*W*, and are fairly consistent across CCT in both color spaces. The present analysis also reveals no notable *overall* relationship between CCT and the differences between IES R_f and CIE R_a for the large set of SPDs (**Figure 10**). This is counter to the conclusion of Khanh, Vin, and Bodrogi [2016c], which was likely influenced by the SPD set. Additionally, $R_{cs,h1}$ values are not CCT-dependent, which reinforces the current findings.

3.2 Influence of CIE R_a and Gamut Shape on Product Development

The influence of metrics on product development can be observed by examining the gamut shape of commercially-available light sources, as was done by Royer, Houser, and David [2017a]. Typical tri-phosphor fluorescent lamps and PC-LED lamps designed to meet the $R_a \geq 80$ criterion increase yellow or yellow/green chroma, because that is the most energy-efficient (and often efficacious) way to meet the average color fidelity criterion. Given the discrepancy in treatment of different hues by IES R_f and CIE R_a, such sources almost all have lower IES R_f values than CIE R_a values. This can be addressed by adjusting the scaling factor used in R_f, as was done by the CIE, but doing so does not address the fact that the rank order of sources varies between IES R_f and CIE R_a due to the gamut shape relationship.

SPDs that increase red chroma versus the reference illuminant generally have lower CIE R_a scores than SPDs that decrease red chroma versus the reference illuminant by the same amount (according to accurate models of human vision). These SPDs have the same IES R_f value. This produces a disincentive for manufacturers to develop sources that increase red chroma if they are using CIE R_a to make engineering decisions. Unfortunately, the same increases in red chroma versus the reference that are discouraged via lower R_a scores have been shown to be preferred in perceptual experiments [Davis and Ohno 2010; Esposito 2016; Ohno and others 2015; Royer and others

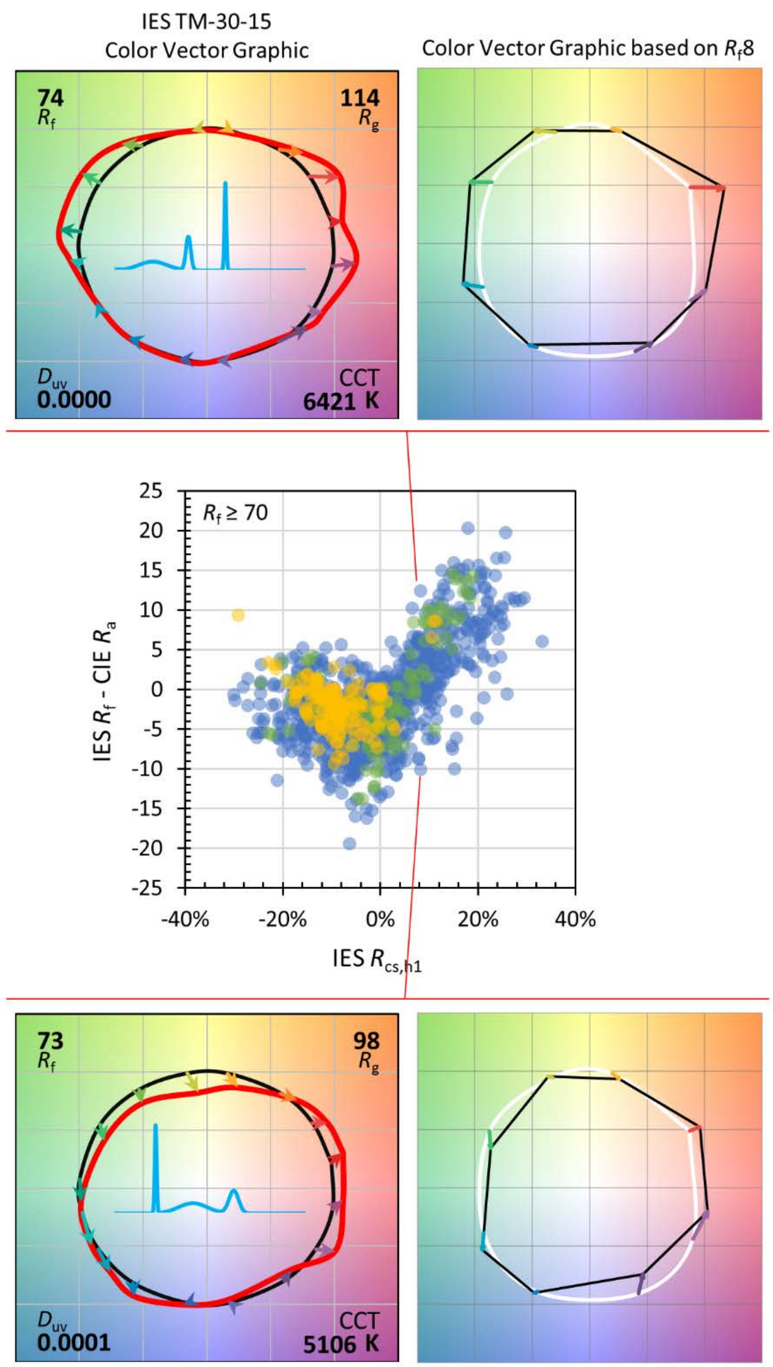

Figure 9. Comparison of example CVGs based on the 99 CES (left) and 8 TCS (right), illustrating the source of some of the variation in the difference between IES R_f and CIE R_a.

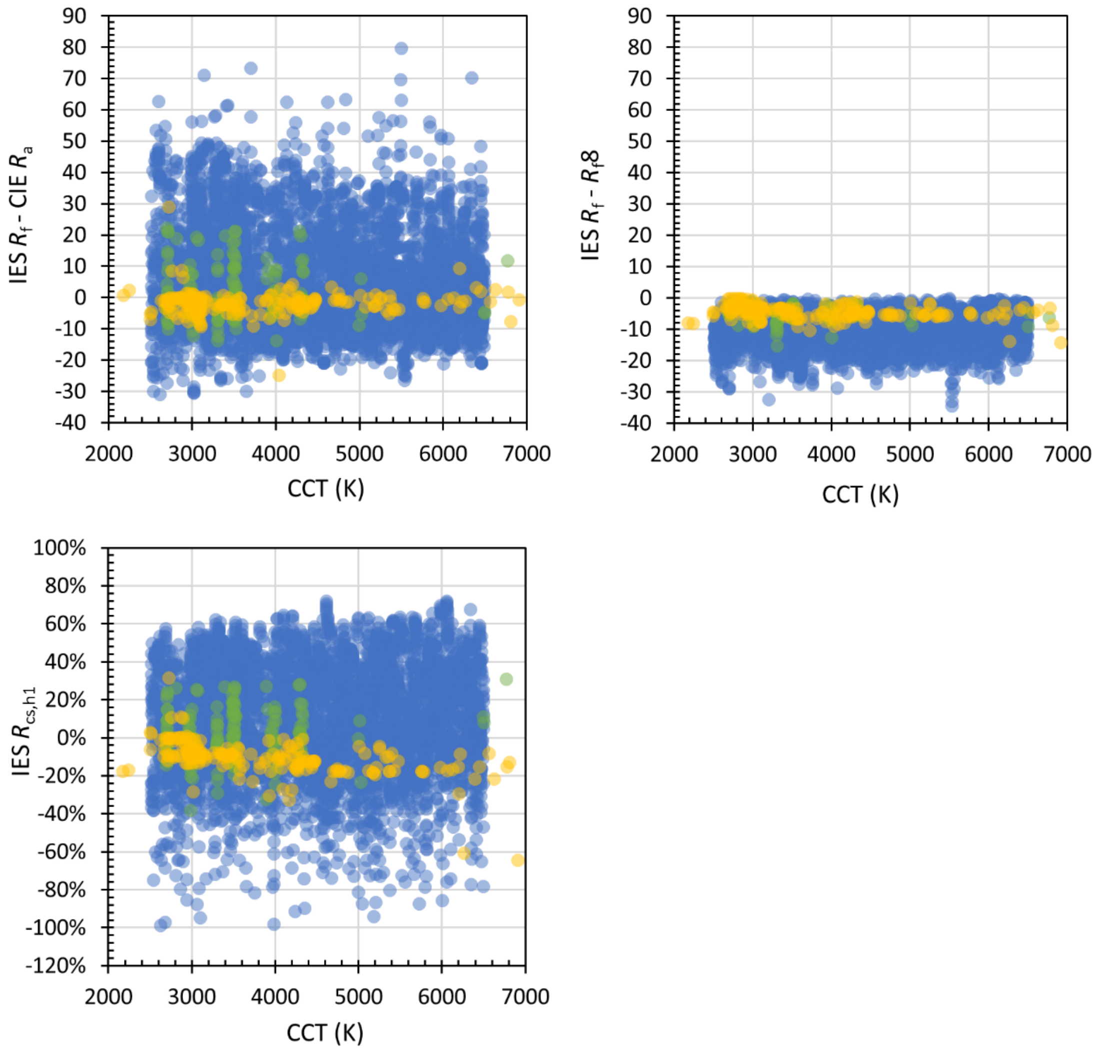

Figure 10. Plots showing the difference between IES R_f and CIE R_a versus CCT (upper left), the difference between IES Rf and R_f8 versus CCT (upper right), and IES $R_{cs,h1}$ versus CCT (bottom). None of the variables plotted are related to CCT.

2017b; Royer and others 2016; Wei and others 2016a; 2016b].

One of the few historical examples of a red-chroma-enhancing light source is the neodymium incandescent lamp. It has a CIE R_a value of 78 and an IES R_f value of 84. The commercial success of the neodymium incandescent lamp [Freeman 2003] likely stems from the fact that consumers do not understand CIE R_a, nor was it frequently reported on consumer packaging until the introduction of LEDs. If the lighting industry's reliance on CIE R_a as a single color rendition measure continues, it will be difficult to market light sources with preferable color rendition. Only with accurate measures of average color fidelity and more comprehensive ways to characterize color rendition that include descriptions of the directions of chroma shifts can the performance of light sources be communicated effectively.

3.3 Accuracy of IES R_f and CIE R_a

Having demonstrated that IES R_f and CIE R_a values are systematically different over a large set of SPDs, a next logical consideration is the accuracy of each as a measure of average color difference—which is what both are intended to be. This is perhaps best examined by considering the key elements of any measure of average color fidelity: the color samples, the reference illuminant, and the color space/chromatic adaptation transformation. It can also be evaluated experimentally.

That there can be "accuracy" associated with test color samples for a standardized measure of color rendition is a misconception. A sample set's "accuracy" is really how well it represents, in terms of distribution in color space and wavelength sensitivity, the object colors (specifically their spectral reflectance functions) of a given architectural space or other environment. Of course, this varies with the environment, and ignores other factors such as the spatial distribution of objects, color contrast, or object importance. Nonetheless, standardized sample sets are necessary for commerce. The rationale for selecting the 99 CES of IES TM-30-15 is thoroughly documented by David and colleagues [2015], offering a strong reason to favor the 99 CES over the 8 TCS in almost all lighting applications. Specifically, the wavelength uniformity and the distribution throughout the color volume of the 99 CES make it a reasonable baseline for predicting performance in a wide variety of spaces.

Likewise, the reference illuminant is a somewhat arbitrary choice, but one that should have some underlying rationale. The reference illuminant is most important if only average color fidelity is considered, but becomes less important if other aspects of color rendition that are not maximized by matching the reference are also considered. The reference illuminants used in IES TM-30-15 differ from those used for CIE R_a only within the CCT range of 4501 K to 5499 K, and were thoroughly examined by Royer [2016].

Finally, the model of human vision—comprising the color space and chromatic adaption transformation—could be considered the strongest influence on the objective accuracy of a measure of average color fidelity. To this end, CAM02-UCS used in IES TM-30-15 has been shown to be more perceptually accurate than other color spaces [Luo and others 2006], including specifically in the context of color rendition [Xu and others 2016]. There is no evidence related to perceptual accuracy to support the use of CIE U*V*W*, the color space used to calculate CIE R_a. This suggests that the biases introduced by CIE R_a are not related to psychophysics and should be considered errors.

3.4 Implications for Experiments Correlating Color Fidelity with Subjective Evaluations

Despite clear evidence that IES R_f (or CIE R_f) is a more scientifically accurate measure of color rendition compared to CIE R_a, there remains a notion that CIE R_a has some value beyond its intended use as a measure of color fidelity [CIE 2017]. This perhaps stems from psychophysical experiments, many of which have been conducted over the past half century. However, many of these studies must now be questioned because more complete characterizations of color rendition redefine a key part of the independent variable.

Especially in the past decade, numerous experiments have sought to correlate human subjective evaluation with CIE R_a or other measures of color rendition [Esposito 2016; Islam and others 2013; Jerome 1972; Jost-Boissard and others 2014; Jost-Boissard and others 2009; Khanh and Bodrogi 2016; Khanh and others 2016a; 2016b; Lin and others 2015; Liu and others 2013; Ohno and others 2015; Quellman and Boyce 2002; Rea and Freyssinier-Nova 2008; Rea and Freyssinier 2010;

Royer and others 2017b; Royer and others 2016; Schanda and Sandor 2003; Smet and others 2010; Spaulding 2012; Szabó and others 2009; Szabo and others 2014; Teunissen and others 2016; Thornton 1974; Veitch and others 2002; Veitch and others 2014; Wadhwa and Davis 1998; Wang and Wei 2017; Wei and others 2014a; 2016a; 2016b; Wei and others 2014b; Xu and others 2016; Zukauskas and others 2012]. A majority of these studies relied on only a small number of SPDs (8 or less) [Jerome 1972; Jost-Boissard and others 2009; Khanh and Bodrogi 2016; Khanh and others 2016a; 2016b; Lin and others 2015; Quellman and Boyce 2002; Rea and Freyssinier-Nova 2008; Rea and Freyssinier 2010; Schanda and Sandor 2003; Smet and others 2010; Spaulding 2012; Szabó and others 2009; Teunissen and others 2016; Thornton 1974; Veitch and others 2002; Veitch and others 2014; Wadhwa and Davis 1998; Wei and others 2014a; Wei and others 2014b], and none included gamut shape as a variable until a handful of recent efforts [Esposito 2016; Royer and others 2017b; Royer and others 2016; Wei and others 2016a; 2016b]. Without sufficient variation in gamut shape, it is possible to find correlations between average fidelity and human subjective evaluations that should not be extrapolated beyond the experimental conditions. To illustrate this possibility, the set of 26 SPDs used by Royer and colleagues [2016] was analyzed. With the complete dataset—which includes multiple SPDs at the same average color fidelity levels having different gamut areas and gamut shapes—it can be shown that neither CIE R_a nor IES R_f is a strong predictor of any of the qualities studied: normalness, saturation, or preference. In fact, CIE R_a has a negative correlation with mean rated preference (**Figure 11**). In contrast, if only the eight SPDs with $R_{cs,h1} \leq 0\%$ are included, which represent the type of red desaturating sources commonly available commercially, both IES R_f ($r^2 = 0.84$) and CIE R_a ($r^2 = 0.74$) have strong positive correlations with mean rated preference. This clearly demonstrates the potential for misleading results if a small number of SPDs that do not consider gamut shape are used in psychophysical experiments.

A large number of combinations of color shifts that can lead to any given color fidelity value. Future experiments seeking to compare the performance of measures of color rendition should consider all aspects of a light source's effect on color appearance—including color rendition, chromaticity, and illuminance. While the growing complexity of characterizing these effects may preclude examining all simultaneously, the independent variable needs to be carefully defined and other factors held constant.

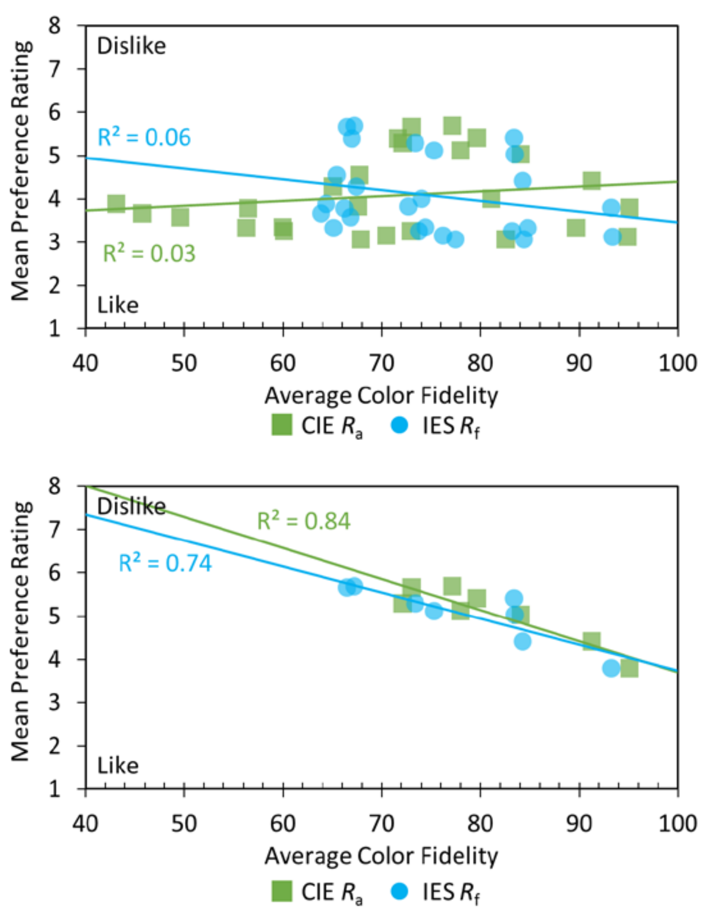

Figure 11. Royer and colleagues [2016] demonstrated that when average color fidelity is balanced with gamut area and gamut shape, average color fidelity is not correlated with rated preference (top). However, if only one gamut shape is considered (here, when $R_{cs,h1}$<0%), average color fidelity may be correlated with rated preference (bottom).

3.5 Is Average Color Fidelity a Useful Tool?

CIE 224:2017, *Color Fidelity Index for Accurate Scientific Use*, states: "…it is considered that such unintended uses of CRI [CIE R_a] as an overall colour quality measure for end users is not better fulfilled by the more scientifically accurate general colour fidelity index, R_f. This is because the users' evaluation is influenced by factors beyond colour fidelity such as chroma effects, and the detailed nature of specific illumination tasks" [CIE 2017]. Said another way, CIE R_f is a scientifically better measure of average color fidelity than CIE R_a, but average color fidelity in general may have limited utility, no matter the measure. Even CIE 13.3-1995 [CIE 1995], which defines CIE R_a and the rest of the CIE Test Color Method, states: "In order to describe fully the colour rendering properties of a light source a series of Special Colour Rendering Indices is necessary…The importance of the directions of the colour shifts is recognized but not included in the Colour Rendering Indices." The critically important—but often overlooked in practice—limitations of average color fidelity as a characterization of color rendition have been documented by Houser and colleagues [2016]. While the work of Houser and colleagues preceded the publication of IES TM-30-15 and is therefore focused on CIE R_a, the same philosophical limitations apply to IES R_f (and CIE R_f).

Because different combinations of color shifts can lead to the same average color fidelity value (and gamut area), measures of average color fidelity are most useful when the intent is to match the color appearance produced by the reference illuminant, and a very high color fidelity value can be achieved (for example, CIE $R_f \geq 95$). This merit may be amplified by using non-standardized average color fidelity calculations that are customized to compare performance to a defined reference illuminant (instead of a standard reference illuminant) and/or customized color samples, as was done for the relighting of the Sistine Chapel [Schanda and others 2016]. In other cases, average color fidelity alone reveals little about how the colors of objects in a space will appear. Once IES R_f drops to a level in the low 80s—typical of many architectural lighting products—or below, two light sources with the same IES R_f (and IES R_g) values may be perceived differently in terms of normalness, saturation, or preference [Royer and others 2016]. In this way, average color fidelity has little value as a tool for predicting color appearance in an architectural space, which is a primary purpose of color rendition measures.

At the same time, pairing average color fidelity with a measure of red chroma shift (but not necessarily gamut area, as suggested by some work [Houser and others 2013; Khanh and others 2016c; Rea and Freyssinier-Nova 2008; Rea and Freyssinier 2010; Teunissen and others 2016]) has been shown to provide useful information for predicting human subjective evaluations of color quality, which are frequently referred to as perceptions [Esposito 2016; Royer and others 2017a; Royer and others 2017b]. This presents the challenge of needing to move away from relying on average color fidelity as a catch-all for color quality, while maintaining the use of average color fidelity as part of a more complete color quality specification.

4 Conclusions

When comparing the values of measures of color rendition against one another, it is critical to use a set of SPDs with varied properties—particularly different gamut shapes. At present, this will require using experimental or theoretical SPDs, because the diversity of gamut shapes for commercially-available products is limited. Given an appropriate set of SPDs for completing a comparison, it is also important to report appropriate statistics for quantifying the difference. The similarity of values for different measures of average color fidelity cannot be described only based on the coefficient of determination (r^2) or coefficient of correlation (r); the analysis should include the range of possible differences (at a particular baseline).

With these two points in mind, a comparison was made between IES R_f and CIE R_a. For a set of 4,945 commercially-available, experimental, and theoretical SPDs, the IES R_f values ranged from 50 to 86 at a CIE R_a value of 80. SPDs that increase red chroma (for example, $R_{cs,h1}$ or $R_{cs,h2}$) tend to have lower CIE R_a values than CIE R_f values due to non-uniformity in the CIE U*V*W* color space. Past comparisons of CIE R_a and IES R_f that relied predominantly or exclusively on SPDs from commercially-available products do not present a complete picture of the difference between IES R_f and CIE R_a. The differences are large and systematic, such that IES R_f provides a meaningful benefit compared to CIE R_a.

This article also demonstrates that it is critical that gamut area and gamut shape be varied along with color fidelity in psychophysical experiments exploring the relationships between human subjective evaluations and measures of color rendition. Otherwise, it is possible—even likely—that the data cannot be appropriately extrapolated beyond the limited extent of the evaluated SPDs.

While acknowledging its weakness, the CIE has not yet deprecated the use of CIE R_a, even with the availability of the greatly superior average color fidelity measure, R_f. Especially in light of the new evidence provided here, the CIE should do so. It is now clear that, despite limited studies that claimed otherwise, the inaccuracies of CIE R_a as a measure of average color fidelity do not make it a stronger correlate of human subjective evaluations of color rendition. In fact, by more strongly penalizing increases in red chroma, CIE R_a discourages development of light sources that people may evaluate as providing more natural, normal, saturated, and/or preferable subjective viewing experiences. Even though CIE R_a is widely used in the lighting industry, it is an impediment to better lighting, so it should be phased out. Of course, while replacing CIE R_a with R_f is necessary, this improvement alone will not be sufficient to enable proper evaluation of color rendition. This will require consideration of other important aspects, including gamut shape and gamut area.

References

[CIE] Commission Internationale de l'Eclairage. 1995. Method of measuring and specifying colour rendering properties of light sources. 3rd ed. Vienna (Austria): CIE. 13.3. 16 p.

[CIE] Commission Internationale de l'Eclairage. 2017. Colour Fidelity Index for accurate scientific use. Vienna, Austra: CIE. 224:2017. 44 p.

David A, Fini PT, Houser KW, Ohno Y, Royer MP, Smet KA, Wei M, Whitehead L. 2015. Development of the IES method for evaluating the color rendition of light sources. Opt Express. 23(12):15888-15906.

Davis W, Ohno Y. 2010. Color quality scale. Optical Eng. 49(3):033602 1-16.

de Beer E, van der Burgt P, van Kemenade J. 2015. Another Color Rendering Metric: Do We Really Need It, Can We Live without It? LEUKOS. 12(1-2):51-59.

Esposito T. 2016. Modeling color rendition and color discrimination with average fidelity, average gamut, and gamut shape [dissertation]. University Park (PA): Penn State University. 176 p.

Freeman K. GE Celebrates Sale of the One Hundred Millionth Reveal Bulb; Strong Marketing Campaign Enhances Sales of 100 Million Bulbs in Less Than Two Years [Internet]. BusinessWire. Available from: http://www.businesswire.com/news/home/20030305005109/en/GE-Celebrates-Sale-Millionth-Reveal-Bulb-Strong

Houser K, Mossman M, Smet K, Whitehead L. 2016. Tutorial: Color Rendering and Its Applications in Lighting. LEUKOS. 12(1-2):7-26.

Houser KW, Wei M, David A, Krames MR, Shen XS. 2013. Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. Opt Express. 21(8):10393-103411.

[IES] Illuminating Engineering Society of North America. 2015. Method for Evaluating Light Source Color Rendition. New York, NY: IES. IES TM-30-15. 26 p.

Islam MS, Dangol R, Hyvarinen M, Bhusal P, Puolakka M, Halonen L. 2013. User preferences for LED lighting in terms of light spectrum. Light Res Technol. 45(6):641-665.

Jerome CW. 1972. Flattery vs Color Rendition. JIES. 1(3):208-211.

Jost-Boissard S, Avouac P, Fontoynont M. 2014. Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference. Light Res Technol. 47(7):769-794.

Jost-Boissard S, Fontoynont M, Blanc-Gonnet J. 2009. Perceived lighting quality of LED sources for the presentation of fruit and vegetables. J Mod Optic. 56(13):1420-1432.

Khanh T, Bodrogi P. 2016. Colour preference, naturalness, vividness and colour quality metrics, Part 3: Experiments with makeup products and analysis of the complete warm white dataset. Light Res Technol. Online Before Print. DOI: 10.1177/1477153516669558.

Khanh T, Bodrogi P, Vinh Q, Stojanovic D. 2016a. Colour preference, naturalness, vividness and colour quality metrics, Part 1: Experiments in a room. Light Res Technol. Online Before Print: DOI: 10.1177/1477153516643359.

Khanh T, Bodrogi P, Vinh Q, Stojanovic D. 2016b. Colour preference, naturalness, vividness and colour quality metrics, Part 2: Experiments in a viewing booth and analysis of the combined dataset. Light Res Technol. Online Before Print: DOI: 10.1177/1477153516643570.

Khanh TQ, Vinh QT, Bodrogi P. 2016c. A numerical analysis of recent colour rendition metrics. Light Res Technol. Online Before Print: DOI: 10.1177/1477153516632568.

Lin Y, Wei M, Smet K, Tsukitani A, Bodrogi P, Khanh T. 2015. Colour preference varies with lighting application. Light Res Technol. 49(3)316-328.

Liu A, Tuzikas A, Zukauskas A, Vaicekauskas R, Vitta P, Shur M. 2013. Cultural Preferences to Color Quality of Illumination of Different Artwork Objects Revealed by a Color Rendition Engine. IEEE Photonics J. 5(4):6801010.

Luo MR, Cui G, Li C. 2006. Uniform colour spaces based on CIECAM02 colour appearance model. Col Res Appl. 31(4):320-330.

Ohno Y, Fein G, Miller C. Vision Experiment on chroma saturation for color quality preference. In: 28th CIE Session; 2015 Jun 28 - Jul 4; Manchester, UK: Vienna (Austria): Commision Internationale de l'Eclairage. 2124 p. 1-10.

Quellman EM, Boyce PR. 2002. The Light Source Color Preferences of People of Different Skin Tones. JIES. 31(1):109-118.

Rea M, Freyssinier JP. 2015. TM-30: Curb Your Enthusiasm. LD+A. New York, NY.: Illuminating Engineering Society. 11-12.

Rea MS, Freyssinier-Nova JP. 2008. Color rendering: A tale of two metrics. Col Res Appl. 33(3):192-202.

Rea MS, Freyssinier JP. 2010. Color Rendering: Beyond Pride and Prejudice. Col Res Appl. 35(6):401-409.

Royer M, Houser K, David A. 2017a. Chroma Shift and Gamut Shape: Going Beyond Average Color Fidelity and Gamut Area. LEUKOS. Online Before Print. DOI:10.1080/15502724.2017.1372203.

Royer M, Wei M. 2017. The Role of Presented Objects in Deriving Color Preference Criteria from Psychophysical Studies. LEUKOS. 13(3):143-157.

Royer M, Wilkerson A, Wei M. 2017b. Human Perceptions of Color Rendition at Different Chromaticities. Light Res Technol. Online Before Print. DOI: 10.1177/1477153517725974.

Royer M, Wilkerson A, Wei M, Houser K, Davis R. 2016. Human perceptions of colour rendition vary with average fidelity, average gamut, and gamut shape. Light Res Technol. Online Before Print. DOI: 10.1177/1477153516663615.

Royer MP. 2016. What is the Reference? An Examination of Alternatives to the Reference Sources Used in IES TM-30-15. LEUKOS. 13(2):71-89.

Schanda J, Csuti P, Szabo F. 2016. A New Concept of Color Fidelity for Museum Lighting: Based on an Experiment in the Sistine Chapel. LEUKOS. 12(1-2):71-77.

Schanda J, Sandor N. 2003. Colour rendering, Past – Present – Future. International Lighting and Colour Conference. Cape Town, South Africa. 76-85.

Smet K, David A, Whitehead L. 2015. Why Color Space Uniformity and Sample Set Spectral Uniformity Are Essential for Color Rendering Measures. Leukos 12(1-2):39-50.

Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. 2010. Memory colours and colour quality evaluation of conventional and solid-state lamps. Opt Express. 18(25):26229-44.

Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. 2011. Correlation between color quality metric predictions and visual appreciation of light sources. Opt Express. 19(9):8151-66.

Smet K, Schanda J, Whitehead L, Luo RM. 2013. CRI2012: A proposal for updating the CIE colour rendering index. Light Res Technol. 45(6):689-709.

Spaulding JM. Evaluation of Desirability Assessment Techniques for Tunable Solid State Lighting Applications. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 22-26 Oct 2012. 643-647.

Szabó F, Csuti P, Schanda J. Color preference under different illuminants—new approach of light source coloour quality. Light and Lighting Conference with Special Emphasis on LEDs and Solid State Lighting; 27-29 May 2009 2009; Budapest, Hungary: CIE. PWDAS-43.

Szabo F, Keri R, Schanda J, Csuti P, Mihalyko-Orban E. 2014. A study of preferred colour rendering of light sources: Home lighting. Light Res Technol. 48(2):103-125.

Teunissen C, van der Heijden F, Poort SHM, de Beer E. 2016. Characterising user preference for white LED light sources with CIE colour rendering index combined with a relative gamut area index. Light Res Technol. 49(4):461-480.

Thornton WA. 1974. A Validation of the Color-Preference Index. Journal of the Illuminating Engineering Society 4(1):48-52.

van der Burgt P, van Kemenade J. 2010. About color rendition of light sources: The balance between simplicity and accuracy. Color Research & Application: 35(2)85-93.

Veitch JA, Tiller DK, Pasini I, Arsenault CD, Jaekel RR, Svec JM. 2002. The effects of fluorescent lighting filters on skin appearance and visual performance. JIES. 31(1):40-60.

Veitch JA, Whitehead LA, Mossman M, Pilditch TD. 2014. Chromaticity-Matched but Spectrally Different Light Source Effects on Simple and Complex Color Judgments. Col Res Appl. 39(3):263-274.

Wadhwa AM, Davis RG. 1998. Lighting of Historical Facades - A Comparison of Different Lamp Technologies. JIES. 27(2):43-58.

Wang Y, Wei M. 2017. Preference among light sources with different Duv but similar colour rendition: A pilot study. Light Res Technol. Online Before Print. DOI: 10.1177/1477153517712552.

Wei M, Houser K, David A, Krames M. 2014a. Perceptual responses to LED illumination with colour rendering indices of 85 and 97. Light Res Technol. 47(7):810-827.

Wei M, Houser K, David A, Krames M. 2016a. Colour gamut size and shape influence colour preference. Light Res Technol. Online Before Print. DOI: 10.1177/1477153516651472.

Wei M, Houser K, David A, Krames M. 2016b. Effect of gamut shape on color preference. CIE 2016 Lighting Quality and Energy Efficiency. Melbourne, Australia: Vienna (Austria): Commission Internationale de l'Eclairage. 32-41.

Wei M, Houser KW. 2012. Status of Solid-State Lighting Based on Entries to the 2010 US DOE Next Generation Luminaire Competition. LEUKOS. 8(4):237-259.

Wei M, Houser KW, Allen GR, Beers WW. 2014b. Color Preference under LEDs with Diminished Yellow Emission. LEUKOS. 10(3):119-131.

Xu W, Wei M, Smet K, Lin Y. 2016. The prediction of perceived colour differences by colour fidelity metrics. Light Res Technol. Online before print. DOI: 10.1177/1477153516653650.

Zukauskas A, Vaicekauskas R, Vitta P, Tuzikas A, Petrulis A, Shur M. 2012. Color rendition engine. Opt Express. 20(5):5356-5367.