Designation: E 2214 – 02

Standard Practice for
Specifying and Verifying the Performance of Color-Measuring Instruments[1]

This standard is issued under the fixed designation E 2214; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

Recent advances in optics, electronics and documentary standard have resulted in a proliferation of instruments for the measurement of color and appearance of materials and objects. These instruments possess very good performance but there has been little progress toward standardizing the terminology and procedures to quantify that performance. Therefore, the commercial literature and even some documentary standards are a mass of confusing terms, numbers and specifications that are impossible to compare or interpret.

Two recent papers in the literature, have proposed terms and procedures to standardize the specification, comparison and verification of the level of performance of a color-measuring instrument.[2,3] Following those procedures, those specifications can be compared to product tolerances. This becomes important so that instrument users and instrument makers can agree on how to compare or verify, or both, that their instruments are performing in the field as they were designed and tested in the factory.

1. Scope

1.1 This practice provides standard terms and procedures for describing and characterizing the performance of spectral and filter based instruments designed to measure and compute the colorimetric properties of materials and objects. It does not set the specifications but rather gives the format and process by which specifications can be determined, communicated and verified.

1.2 This practice does not describe methods that are generally applicable to visible-range spectroscopic instruments used for analytical chemistry (UV-VIS spectrophotometers). ASTM Committee E13 on Molecular Spectroscopy and Chromatography includes such procedures in standards under their jurisdiction.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*

D 2244 Practice for Calculation of Color Tolerances and Color Differences from Instrumentally Measured Color Coordinates[4]

E 284 Terminology of Appearance[4]

E 1164 Practice for Obtaining Spectrophotometric Data for Object-Color Evaluation[4]

2.2 *Other Documents:*

ISO International Vocabulary of Basic and General Terms in Metrology (VIM)[5]

NIST Technical Note 1297 Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results[6]

DIN 55600 Bestimmung der Signifikanz von Farbabständen bei Körperfarben nach der CIELAB-Formel[7]

3. Terminology

3.1 Definitions of appearance terms in Terminology E 284 are applicable to this pracitce.

---

[1] This practice is under the jurisdiction of ASTM Committee E12 on Color and Appearance and is the direct responsibility of Subcommittee E12.04 on Color and Appearance Analysis.

Current edition approved June 10, 2002. Published August 2002.

[2] Ladson, J., "Colorimetric Data Comparison of Bench-Top and Portable Instruments," *AIC Interim Meeting, Colorimetry*, Berlin, 1995.

[3] Rich, D., "Standardized Terminology and Procedures for Specifying and Verifying the Performance of Spectrocolorimeters," *AIC Color 97 Kyoto*, Kyoto 1997.

[4] *Annual Book of ASTM Standards*, Vol 06.01.

[5] ISO/IDE/OIML/BIPM, *International Vocabulary of Basic and General Terms in Metrology*, International Organization for Standardization, Geneva, Switzerland, 1984.

[6] Taylor, Barry N., and Kuyatt, Chris E., *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297, U. S. Government Printing Office, Washington, D. C., 1984.

[7] DIN, Deutsches Institut für Normung,-Taschenbuch 49, Farbmittel 1, Pgimente, Füllstoffe, Farbstoffe, DIN 5033 Teil 1 bis DIN 55929, Beuth Verlag GmbH, Berlin.

3.2 Definitions of metrology terms in ISO, International Vocabulary of Basic and General Terms in Metrology (VIM) are applicable to this practice.

3.3 *Definitions of Terms Specific to This Standard:*

3.3.1 *colorimetric spectrometer*, *n*—spectrometer, one component of which is a dispersive element (such as a prism, grating or interference filter or wedge or tunable or discrete series of monochromatic sources), that is normally capable of producing as output colorimetric data (such as tristimulus values and derived color coordinates or indices of appearance attributes) as well as the underlying spectral data from which the colorimetric data are derived.

3.3.1.1 *Discussion*—At one time, UV-VIS analytical spectrophotometers were used for colorimetric measurements. Today, while instruments intended for use in color measurements share many common components with UV-VIS analytical spectrometers, there are two distinct classes of instruments. UV-VIS analytical spectrometers are designed to optimize their use in chemometric quantitative analysis, which requires very precise spectral position and very narrow spectral windows and moderate baseline stability. Colorimetric spectrometers are designed to optimize their use as simulations of the visual colorimeter or as the source of spectral and colorimetric information for computer-assisted color matching systems. They allow more tolerance on the spectral scale and spectral window width but demand much more stability in the radiometric scale.

3.3.2 *inter-instrument agreement*, *n*—a form of reproducibility in which two or more instruments from the same manufacturer and model are compared.

3.3.3 *inter-model agreement*, *n*—a form of reproducibility in which the measurements of two or more instruments from different manufacturers or of different but equivalent design are compared.

3.3.3.1 *Discussion*—Modern instruments have such high precision that small differences in geometric and spectral design can result in significant differences in the performance of two instruments. This can occur even though both instruments exhibit design and performance bias which are well within the expected combined uncertainty of the instrument and within the requirements of any international standard.

3.3.4 *mean color difference from the mean, MCDM, n*—a measure of expectation value of the performance of a color-measuring instrument.

3.3.4.1 *Discussion*—MCDM calculates the average color difference between a set of readings and the average of that set of readings. MCDM = average(ΔE_i(average(Lab) − Lab_i)), for $i = 1$ to N readings. Any standard color difference or color tolerance equation can be used as long as the report clearly identifies the equation being used (see Practice D 2244).

4. Summary of Practice

4.1 This practice defines standardized terms for the most common instrument measurement performance parameters (repeatability, reproducibility, inter-instrument agreement, inter-model instrument agreement, accuracy) and describes a set of measurements and artifacts, with which both the producers and users of color-measuring instruments verify or certify the specification and performance of color-measuring instruments. Following this practice can improve communications between instrument manufacturers and instrument users and between suppliers and purchasers of colored materials.

5. Significance and Use

5.1 In today's commerce, instrument makers and instrument users must deal with a large array of bench-top and portable color-measuring instruments, many with different geometric and spectral characteristics. At the same time, manufacturers of colored goods are adopting quality management systems that require periodic verification of the performance of the instruments that are critical to the quality of the final product. The technology involved in optics and electro-optics has progressed greatly over the last decade. The result has been a generation of instruments that are both more affordable and higher in performance. What had been a tool for the research laboratory is now available to the retail point of sale, to manufacturing, to design and to corporate communications. New documentary standards have been published that encourage the use of colorimeters, spectrocolorimeters, and colorimetric spetrometers in applications previously dominated by visual expertise or by filter densitometers.[8] Therefore, it is necessary to determine if an instrument is suitable to the application and to verify that an instrument or instruments are working within the required operating parameters.

5.2 This practice provides descriptions of some common instrumental parameters that relate to the way an instrument will contribute to the quality and consistency of the production of colored goods. It also describes some of the material standards required to assess the performance of a color-measuring instrument and suggests some tests and test reports to aid in verifying the performance of the instrument relative to its intended application.

6. Instrument Performance Parameters

6.1 *Repeatability* is generally the most important specification in a color-measuring instrument. Colorimetry is primarily a relative or differential measurement, not an absolute measurement. In colorimetry, there is always a standard and a trial specimen. The standard may be a real physical specimen or it may be a set of theoretical target values. The trial is usually similar to the standard in both appearance and spectral nature. Thus, industrial colorimetry is generally a test of how well the instrument repeats its readings of the same or nearly the same specimen over a period of minutes, hours, days, and weeks.

6.1.1 The ISO VIM defines repeatability as a measure of the random error of a reading and assumes that the sample standard deviation is an estimate of repeatability. Repeatability is further defined as the standard deviation of a set of measurements taken over a specified time period by a single operator, on a single instrument with a single specimen. This definition is similar to that in Terminology E 284, except that the ISO explicitly defines the metric of "closeness of agreement" as the sample standard deviation. Since color is a multidimensional property of a material, repeatability should be reported in terms

---

[8] ISO 13655 Spectral Measurement and Colorimetric Computation for Graphic Arts Images, International Organization for Standardization, Geneva, Switzerland.

of the multidimensional standard deviations, derived from the square root of the absolute value of the variance–covariance matrix.

6.1.2 The time period over which the readings are collected must be specified and is often qualitatively described as "short," "medium," or "long." The definitions of these time frames do not overlap. This is intentional, providing clearly defined milestones in the temporal stability of test results.

6.1.2.1 For the purposes of colorimetry, "short" is normally the time required to collect a set of 30 readings, taken as fast as the instrument will allow. The actual time will vary as a function of lamp and power supply characteristics but should be less than one hour.

6.1.2.2 "Medium" term is normally defined as, at least the period of one work shift (8 h) but less than three work shifts (one day).

6.1.2.3 "Long" term is open ended but is often described as any set readings taken over a period of at least 4 to 8 weeks. The longest known reported study described readings taken over a period of 3¼ years.[9]

6.2 *Reproducibility* is the second most important specification in a color-measuring instrument. According to Terminology E 284, reproducibility is a form of repeatability in which one or more of the measurement parameters have been systematically changed. Thus the sample is different, the procedures or instrument are different, or the time frame is very long. The increase of disorder over a very long time changes the instrument systematically and the set of readings really compares a "young" instrument with an "old" instrument.

6.2.1 The ISO VIM defines reproducibility as a type of repeatability in which either the time frame is very long, in which the operator changes, the instrument changes, or the measurement conditions change. ISO again recommends estimating this with a standard deviation. Reproducibility is further defined as the standard deviation of a set of measurements taken over a specified period of time by a single operator, on a single instrument with a single specimen. This definition is similar to that in Terminology E 284, except that the ISO again, explicitly defines the metric of "closeness of agreement" as the sample standard deviation. Again, since color is a multidimensional property of a material, reproducibility should be reported in terms of the multidimensional standard deviations, derived from the square root of the absolute value of the variance–covariance matrix.

6.2.2 The time period over which the readings are collected must be specified. For the purposes of colorimetry, "long" term repeatability is the most common and important type of reproducibility. Repeatability and reproducibility have traditionally been evaluated in colorimetry by comparing the color differences of a set of readings to a single reading or to the average of the set of readings.

6.3 *Inter-Instrument Agreement*, as defined in 3.3.2, describes the reproducibility between two or more instruments, of identical design. The ISO has no definition or description of such a concept. This is because in most test results, a method or instrument dependent bias can be assessed. In this situation, such a test measures the consistency of the design and manufacturing process. Within the technical description of the standard geometric and spectral parameters for the measurement of diffuse reflectance factor and color, a significant amount of latitude exists. This latitude results in a random amount of bias. For a given design, a manufacturer may reduce the random bias, often to a level less than the stability of reference materials. The most common form of test for inter-model instrument agreement is pairwise color difference assessment of a series of specimens. Various parameters are reported in the literature including the average color difference, the maximum color difference, the typical color difference, the RMS color difference or the MCDM mean color difference from the mean, taking the average of all instruments as the standard and the other as the test instrument. Using pairs of instruments and materials one can derive a multivariate confidence interval against the value 0.0 difference and then test individual components to determine which attribute (lightness, chroma, hue) are the significant contributors to the differences between instruments. If a group of instruments are being tested then a multivariate analysis of variance (MANOVA) can be performed to test the agreement of the means of the instrument.

6.4 *Inter-Model Agreement*, as defined in 3.3.3, describes the reproducibility between two or more instruments of differing design. The latitude within the standard geometric and spectral parameters described in the preceding paragraph is at a maximum when the designs differ. The systematic bias may increase by factors of from 5 to 10 because of the increased latitude. Standardizing laboratories will report either the algebraic differences between measurement results or the ratio of the measurement values between two labs. The former will be a Normal statistical variable if the measurement values are Normally distributed, and the latter will be distributed as a non-central F distribution with the expected value equal to the bias. This non-central distribution can be estimated from the multivariate variance–covariance matrix. Using pairs of instruments and materials one can derive a multivariate confidence interval against the value 0.0 difference and then test individual components to determine which attribute (lightness, chroma, hue) are the significant contributors to the differences between instruments. If a group of instruments are being tested then a multivariate analysis of variance (MANOVA) can be performed to test the agreement of the means of the instrument.

6.5 *Accuracy*, while occasionally critical, is generally the least significant parameter in characterizing the performance of a color-measuring instrument. ISO defines accuracy as the conformance of a series of readings to the accepted or true value. In modern colorimetry, the volume of the total combined uncertainty around the accepted value is often many times larger than volume of visual acceptability of the products whose color is being quantified. Therefore, an "accurate" color measurement may result in an unacceptable product color. There are two scales in a spectrocolorimeter that can be assigned nominal values and tested against standard values. They are the radiometric scale and the wavelength scale.

6.5.1 The wavelength scale includes the sampling position

---

[9] Rich, D. C., Battle, D., Malkin, F., Williamson, C., Ingleson, A., "Evaluation of the Long-Term Repeatability of Reflectance Spectrophotometers," *Spectrophotometry, Luminescence and Colour: Science and Compliance*, C. Burgess and D. G. Jones, eds., Elsevier, Amsterdam, 1995.

(centroid wavelength) and the sampling window width (spectral bandwidth). These parameters are normally tested against physical standards of wavelength based on fundamental phenomena, such as discharge lamps or laser lines. In very abridged instruments it may not be possible to test directly against such a physical standard, so either material standards are used, such as holmium oxide or didymium oxide glasses or pairs of sharp-cutting filter glasses, or a scanning monochromator are characterized against physical standards. In the case of the monochromator, the output intensity is equalized and scanned across the input to the abridged spectrometer to resolve the location of the wavelength centroid at each sampling point in the abridged spectrum.

6.5.2 Radiometric scale accuracy is more difficult to evaluate since it involves three aspects: the zero level, white level, and the linearity between the two levels. White level can be tested by direct comparison to a primary standard of reflectance or transmittance and the result reported as ± the expanded uncertainty at a stated confidence level, as described in NIST Technical Note 1297. The expanded uncertainty is the combined uncertainty of the white plaque and the instrument under test combined in quadrature at the 95 % confidence level and multiplied by the appropriate coverage factor. The exact methods for propagating the uncertainty in a reflectance factor measurement into the color coordinates is still a matter of some dispute. Methods have been proposed in the literature but are not widely accepted and used.[10]

6.5.3 The black level only needs to be tested to show that the optical zero is less than some minimum value, since it is impossible to define the optical zero except in terms of the noise floor of the spectrometer or colorimeter. The results of measurements of near black materials (black glass of known refractive index or a suitably designed black trap) shall show results that are less than some upper limit. For example, the zero level $\leq$ 0.025 %.

6.5.4 Finally, linearity must be specified in a testable way. If the spectrometer is linear then at any wavelength, plots of the measured values versus the standard values of a set of neutral samples should lie on a line passing through the origin with a slope of 1.0. Unfortunately, it is possible to fit a line by least squares to a higher order function (having some errors positive and some negative) and obtain a slope of 1.0. Estimating the slope of the line passing through all points will not identify that kind of non-linearity. To avoid this, standardizing laboratories and some analytical instruments use the addition-of-radiance method, either with two sources or with a double aperture apparatus to generate a signal and a 2× signal into the spectrometer that can be adjusted to cover the radiometric range of the spectrometer. Since commercial colorimeters are not easily configured with such devices, the use of neutral plaques or neutral filters is the best compromise.

7. Procedures

7.1 Repeatability shall be measured by placing a white plaque at the measurement port of a recently standardized instrument and making replicate readings of the plaque without moving the plaque. For short-term repeatability, at least 30 readings shall be collected as fast as the instrument allows. The quantity of reading (30 or more) depends upon the desired level of confidence in the results and the time required to acquire that number of readings. For very slow instruments, the costs of performing even 30 measurements may be very high, in those cases a lower number of readings may be adequate if the variance-covariance is adequately characterized. For medium term repeatability, at least 60 readings shall be collected, uniformly spread out over an 8-h period, with at least 60 s between readings. Use of a white plaque is recommended because the radiometric random noise is generally highest near the upper end of the scale of diffuse reflectance. A noise level of a few hundredths of a percent is expected at a 90 % reflectance while the noise level may be a few thousandths of a percent at 4 % reflectance. Spectrally selective (colored) standards are not recommended as they tend to confound the radiometric noise with temperature and mechanical sensitivity in a way that is not representative of the general performance of the instrument. Often, a light gray plaque may be substituted for the white plaque when an instrument is never used to measure very light or white specimens as the gray level may result in values for repeatability that are more representative of typical materials. Measurements of medium, dark or black specimens will not generally add any useful information since the radiometric noise level tends to be proportional to the signal and the noise will be lost inside the resolution limit of the spectrometer.

7.1.1 The basic measurement values of a colorimetric spectrometer are the set of reflectance factors and those of a filter colorimeter are the tristimulus values. Those variables are the most closely related to a Normal statistical random variable. The reported repeatability shall be either twice the univariate standard deviation of at least three, widely separated reflectance factors, $\Delta R_{\lambda}(2\sigma)$, ($\lambda$ = 440 nm, 560 nm, 660 nm) that are returned from the instrument, or if the reflectance factors are not available, then twice the univariate standard deviation of the individual tristimulus values $\{\Delta X\,(2\sigma), \Delta\,Y\,(2\sigma), \Delta Z\,(2\sigma)\}$. Since the variance (standard deviation) of closely spaced spectral reflectance factors or tristimulus values are generally, not independent, it will be necessary to report the multivariate variance-covariance matrix instead of the square root of the variance for each measurement point. If the set of univariate standard deviations (multiple reflectance factors or multiple tristimulus values) must be reduced to a single dimension, then it is recommended that a weighted standard deviation be reported, the weight being proportional to the sum of the standard observer functions [weight = $(\bar{x}, + \bar{y}, + \bar{z})$] or the variances of the tristimulus values themselves $(\sigma^2_X + \sigma^2_Y + \sigma^2_Z)$.

7.1.1.1 If an estimate of expected color difference is desired then it can be reported either as the average difference from the first readings or as the multivariate confidence volume which will be a three dimensional ellipsoid for tristimulus data and an N dimensional hyper-ellipsoid for N spectral values. Annex A1 shows the steps required to compute the multidimensional confidence interval. This method is documented in DIN 55600

---

[10] Fairchild, M. D., and Reniff, L., "Propagation of Random Errors in Spectrophotometric Colorimetry," *Color Research& Application*, 16, 1991, p. 360.

(in German only) and in two publications in English.[11,12] Reports in the literature have shown that high precision reflectance factor and tristimulus value measurements are seldom distributed as a univariate Normal random variable but are not so far from being distributed as a multivariate Normal random variable.[13] Thus, parameters like variance, standard deviation, root mean square deviations are not appropriate when computed for color difference values.

7.1.1.2 It is also acceptable to report the repeatability in ranges. For example, one value for wavelengths less than 460 nm, a second value for wavelengths between 460 nm and 640 nm and a third value for wavelengths greater than 640 nm, using the appropriate number of wavelengths in the computation of the multivariate confidence volume.

7.2 Reproducibility shall be measured by collecting readings on a set of at least 10 material standards, including both neutral and chromatic samples. It is important to standardize the instrument before each measurement series. Long term repeatability involves daily measurement of the standard for a period of at least 30 days.

NOTE 1—This shall be 30 measurement days, not 30 calendar days.

7.2.1 Reproducibility shall be reported as either two times the univariate standard deviation of the 30 readings of the reflectance factor $\Delta R_\lambda(2\sigma)$ at the same three wavelengths or as twice the univariate standard deviation of the tristimulus values $\{\Delta X\ (2\sigma), \Delta Y(2\sigma), \Delta Z\ (2\sigma)\}$. However, keep in mind that ΔX, ΔY andΔ Z are generally correlated and not independently distributed. Again, to compensate for correlation between tristimulus values or to provide statistics based on uniform color spaces then the multivariate confidence interval shall be computed and reported as described in Annex A1.

7.3 Accuracy has to be independently tested on each of the two scales, wavelength and radiometric. The wavelength scale has the advantage that there are physical standards of wavelength that can be utilized by some colorimetric spectrometers. Procedures for verifying the wavelength scale, bandwidth and radiometric scale are described in Practice E 1164. Specific additional procedures are given here.

7.3.1 The wavelength scale includes the sampling position (centroid wavelength) and the sampling window width (bandwidth). These are normally tested against physical standards of wavelength such as a discharge lamp or laser line. In very abridged instruments it may not be possible to test directly against a physical standard, so one of two options may be exercised. In option 1, a material standard is used, such as holmium oxide or didymium oxide glasses. In option 2, a modern, digital, direct scanning monochromator is characterized against physical sources and the output intensity is equalized and scanned across the input to the abridged spectrometer.

7.3.2 It is recommended that for sampling frequencies of fewer than 16 points across the visible region (400 nm to 700 nm) the wavelength accuracy and bandwidth or filter fit, be tested and reported at each sampling point. Small numbers of spectral samples are usually more independent and have wider spectral windows making each sample point more critical. For sampling frequencies of 16 or more points it is recommended to report the wavelength scale conformance and bandwidth at three specific wavelengths (450 nm ± 0.*x* nm, 550 nm ± 0.*x* nm, 650 nm ± 0.*x* nm) with bandwidths of (*bw* ± 0.*x* nm). Here the tolerances are twice the sample standard deviation for at least 10 replicate determinations of the wavelength scale. If there is a significant bias in the scale position, then that shall be reported as well.

7.3.2.1 As indicated in 6.5.2, the radiometric scale accuracy is more difficult to assess. White level must be tested by direct comparison to a primary standard of reflectance or transmittance obtained from a suitable standardizing laboratory. The result shall be reported along with the expanded uncertainty, as described in 6.5.2.

7.3.2.2 The results of the measurements of the white level can be reported as 100 % ± U %. Using white ceramic plaques with luminous reflectance factors (Y) of 85 or greater, the 100 % level expanded uncertainty will be in the range of ± 0.3 % or the uncertainty of the primary standard, whichever is larger. NPL currently cites uncertainties of ± 0.5 % (2σ) and NIST cites uncertainties of ± 0.3 % (2σ) on white primary standards of reflectance.

7.3.2.3 As indicated in 6.5.3, the black level needs only to be tested to the extent to show that the optical zero is less than some value. The results of measurements of a near black material or a black trap shall be less than 0.0005.

7.3.2.4 As indicated in 6.5.4, the linearity must be evaluated using neutral standards. The five neutral BCRA tiles, the three grays (Pale, Mid, Deep) and the White and Black are recommended for this.[14] The plaques must have standard values assigned to them. Lines are to be passed through each sequential pair of the measured points. The slopes of each line segment shall be compared to the expected value of 1.0. Fig. 1 illustrates that the readings from the five BCRA neutral tiles create four intermediate linear regions. If lines are passed through each of these and the slopes of each compared then differential non-linearities will be seen. Plot or tabulate the results for each line segment: the accepted values for the slope (always 1.0); the measured values for the slopes; and the percent difference between the two. The report shall include the maximum absolute difference ± combined uncertainty in the slope due to uncertainty in the accepted values and the measured values.

7.4 Between Instrument Agreement must report whether the test is for inter-model instrument agreement or inter-instrument agreement. The two types of reproducibility are tested in a similar manner but the results are evaluated quite differently. The difference between these two parameters can be as large as an order of magnitude.

7.4.1 To estimate this reproducibility calculate the mean CIELAB component differences between the readings of a set

---

[11] Jackson, J. E., "Some Multivariate Statistical Techiques Used in Color Matching Data," *J. Opt. Soc. Am*, 49, 1959, pp. 585-592.

[12] Völz, H. G., *Industrial Color Testing Fundamental and Techniques*, VCH VmbH, Weinheim, 1995.

[13] Billmeyer, Jr., F. W., and Alessi, P. J., "Assessment of Color-Measuring Instruments," *Color Research & Application*, 6, 1981, p. 195.

[14] BCRA tiles are produced by CERAM Research, Queens road, Penkhull, Stoke-on-Trent, ST4 7LQ, England.

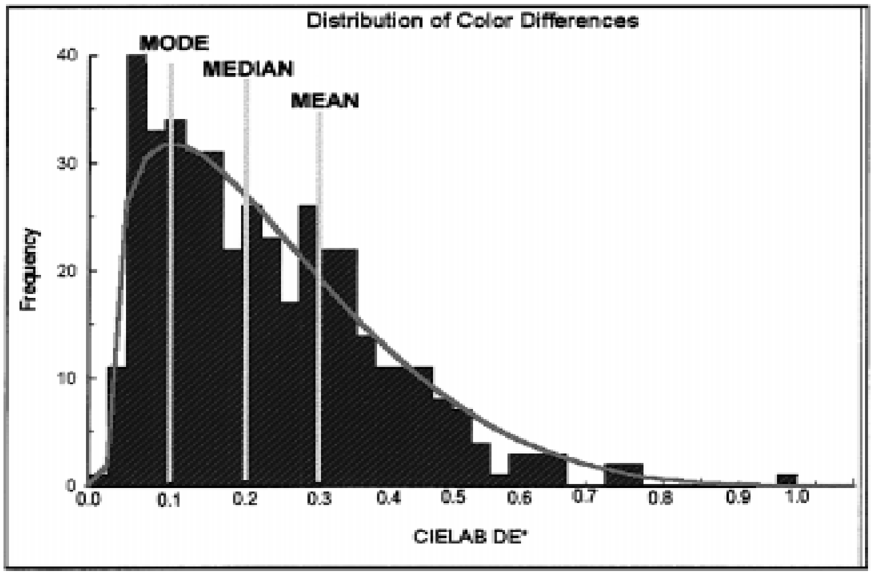
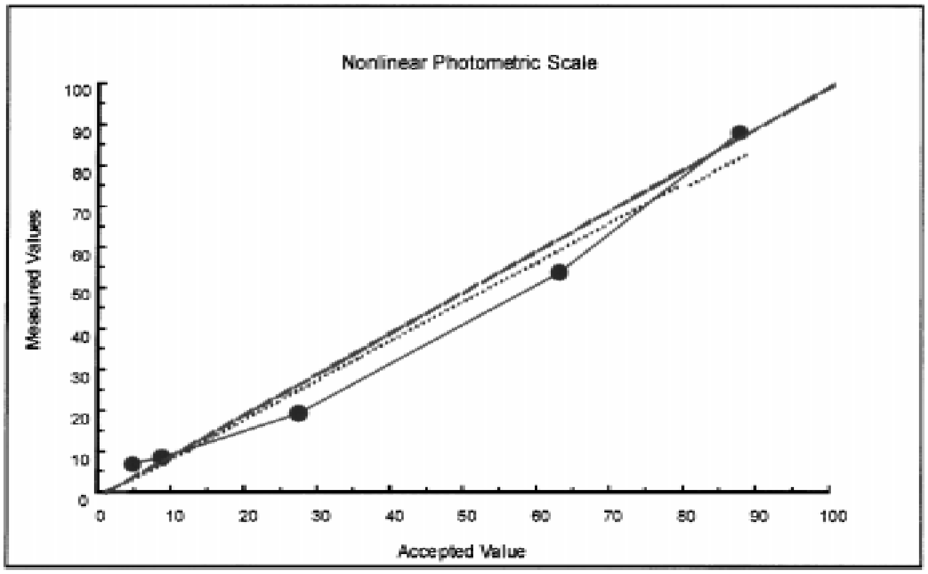


FIG. 1 Comparison of a Large Number of Color Differences Showing Positive Skew Distribution

of at least 10 BCRA tiles plus white and black plaques. The measurement conditions (temperature, humidity, etc.) shall be specified and corrected to standard conditions via the NPL table, reproduced in Annex A2.[15] As observed in Fig. 2, the mode, median and mean of a set of color difference (ΔE) determinations do not follow a bell curve but a curve related to the Chi-squared or F statistical distributions, as described by Hotelling.[16]



NOTE 1—Solid line is ideal, dotted line show zero and a scale error, and points show nonlinear scale.

FIG. 2 Nonlinear Photometric Scale

7.4.1.1 For this positively skewed distribution, if the univariate MODE is reported, then the estimate is highly optimistic, influenced by the easily characterized neutral samples. On the other hand, if the univariate MEDIAN is used then fully half of the readings are above and half below the modal color difference. Finally, the mean (arithmetic average) is highly influenced by the largest color differences in the tail of the distribution. The maximum of the readings is highly dependent on the sample character and the measurement conditions—the maximum will be large for difficult to measure specimens and small for easily characterized specimens.

7.4.2 Spectral ratios or spectral differences can be used to quantify the spectral and radiometric differences between instruments, but geometric differences are confounded with the radiometric differences and thus provide additional information

---

[15] Verrill, J. F., Compton, J. A., and Malkin, F., *Applied Optics*, 25, 1986, p. 3011.

[16] Hotelling, H., "The Generalization of the Student's Ratio," *Annals of Mathematical Statistics*, 2, 1931, pp. 360-378.

only for inter-model instrument agreement or for comparison of spectrally non-selective specimens.[17]

7.4.3 The most appropriate and conservative estimate of the expected difference between the two instruments is the 3D ellipsoidal confidence interval on the joint distribution of measurements from the two instruments. If an estimate of the dispersion of the color differences is required, then compute the variance-covariance matrix and report this or the individual component contributions are described in Annex A1. Replicate readings at different times will mingle the repeatability with the reproducibility and can provide an even better estimate of the expected performance of the instruments under test. The test can be extended to multiple instruments with complete generality, creating a multivariate analysis of variance (MANOVA) problem. The solution of that comparison is beyond the scope of this standard but can be found in most elementary textbooks on multivariate analysis and in any good statistical software package.

8. Keywords

8.1 color; instrumental measurement-color/light; inter-instrument; inter-model; light-transmission and reflection; reflectance; reflectance and reflectivity; spectrocolorimetry; spectrophotometry

ANNEXES

(Mandatory Information)

A1. MULTIVARIATE ANALYSIS OF COLOR READINGS AFTER DIN 55600

A1.1 Estimates of the Mean Vector μ and Variance Matrix V

A1.1.1 Given a set of n color readings, either repeatability or reproducibility readings of a single specimen expressed in terms of CIELAB color coordinates (L^*,a^*,b^*), compute the following:

$$\overline{L^*} = \frac{1}{n}\sum L_i^*,\ \overline{a^*} = \frac{1}{n}\sum a_i^*,\ \overline{b^*} = \frac{1}{n}\sum b_i^* \qquad \text{(A1.1)}$$

$$var(L^*) = \frac{1}{n-1}\sum(L_i^* - \overline{L^*})^2$$

$$var(a^*) = \frac{1}{n-1}\sum(a_i^* - \overline{a^*})^2 \qquad \text{(A1.2)}$$

$$var(b^*) = \frac{1}{n-1}\sum(b_i^* - \overline{b^*})^2 \qquad \text{(A1.2)}$$

$$\mathrm{cov}\ ar(L^*, a^*) = \frac{1}{n-1}\sum(L_i^* - \overline{L^*})\cdot(a_i^* - \overline{a^*}) \qquad \text{(A1.2)}$$

$$\mathrm{cov}\ ar(L^*, b^*) = \frac{1}{n-1}\sum(L_i^* - \overline{L^*})\cdot(b_i^* - \overline{b^*}) \qquad \text{(A1.3)}$$

$$\mathrm{cov}\ ar(a^*, b^*) = \frac{1}{n-1}\sum(a_i^* - \overline{a^*})\cdot(b_i^* - \overline{b^*}) \qquad \text{(A1.3)}$$

A1.1.2 The mean vector μ is given by:

$$i = \begin{bmatrix} \overline{L^*} \\ \overline{a^*} \\ \overline{b^*} \end{bmatrix} \qquad \text{(A1.4)}$$

A1.1.3 The variance–covariance matrix is given by:

$$V = \begin{bmatrix} var(a^*) & cov\ ar(a^*, b^*) & cov\ ar(L^*, a^*) \\ cov\ ar(a^*,b^*) & var(b^*) & cov\ ar(L^*, b^*) \\ cov\ ar(L^*, a^*) & cov\ ar(L^*, b^*) & var(L^*) \end{bmatrix} \qquad \text{(A1.5)}$$

A1.1.4 The variance–covariance matrix must be inverted to form the metric of the state space defined by the measurement variables. That matrix is given the symbol G and has elements $g_{i,j}$, which are similar to the coefficients derived by MacAdam for his ellipses and ellispoids.

$$G = V^{-1} = \begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} \\ g_{2,1} & g_{2,2} & g_{2,3} \\ g_{3,1} & g_{3,2} & g_{3,3} \end{bmatrix} \qquad \text{(A1.6)}$$

A1.2 Application of Multivariate Analysis to Colorimetry

A1.2.1 Critical values for test statistics and for confidence intervals can be derived from the mean vector (μ) and the inverse of the variance-covariance matrix G. The mean vector and variance matrix can be derived for a set individual readings, as shown above, or for a set of color difference readings (Batch–Standard) and critical tests applied. For example, one might want to state that the medium term repeatability is less than 0.05 CIELAB ΔE^* units. Then a series of 30 readings would be subtracted from the initial reading and submitted to this analysis. The mean vector would then be:

$$i = \begin{bmatrix} \overline{\Delta a^*} \\ \overline{\Delta b^*} \\ \overline{\Delta L^*} \end{bmatrix} \qquad \text{(A1.7)}$$

A1.2.2 The critical values are related to the Chi-Squared distribution (χ^2, where critical values for that distribution are for 1−α probability are (70 % = 3.665; 95 % = 7.81; 99 %=11.34). For the color coordinates, the following tables give the formulas for one-sided tests of the absolute values of the color difference components. The critical values for testing is to compare the average values of Δa^*, Δb^*, ΔL^*, ΔC^*, ΔE^*_{ab} to the values in the third column of Table A1.1.

---

[17] Robertson, A. R., *Advances in Standards and Methodology in Spectrophotometry*, Burgess, C., and Mielenz, K. D., eds., Elsevier, Amsterdam, 1987, pp. 277-286.

TABLE A1.1 Critical Values for Testing Significance of Color Differences

Color Coordinate	Critical Value	Equations
Δa^*	$t_{\Delta a}$	$\sqrt{\frac{\chi^2}{n \cdot g_{1,1}}}$
Δb^*	$t_{\Delta b}$	$\sqrt{\frac{\chi^2}{n \cdot g_{2,2}}}$
ΔL^*	$t_{\Delta L}$	$\sqrt{\frac{\chi^2}{n \cdot g_{3,3}}}$
ΔH^*_{ab}	$t_{\Delta H}$	$f = \frac{1}{2}(\overline{h_{Trial}} - \overline{h_{Standard}})$ $g_H = g_{1,1} \sin^2 f - 2g_{12} \sin f \cos f + g_{22} \cos^2 f$ $\sqrt{\frac{\chi^2}{n \cdot g_H}}$
Δh^*_{ab}	$t_{\Delta h}$	$\arctan\left(2 \cdot \frac{t_{\Delta H}}{\overline{\Delta H^*_{ab}}} \cdot \tan\frac{\overline{\Delta h^*_{ab}}}{2}\right)$
ΔC^*_{ab}	$t_{\Delta C}$	$g_{Trial} = g_{1,1} \cos^2\overline{h_{Trial}} + 2g_{1,2} \sin\overline{h_{Trial}} \cos\overline{h_{Trial}} + g_{2,2} \sin^2\overline{h_{Trial}}$ $g_{Std} = g_{1,1} \cos^2\overline{h_{Std}} + 2g_{1,2} \sin\overline{h_{Std}} \cos\overline{h_{Std}} + g_{2,2} \sin^2\overline{h_{Std}}$ $\sqrt{\frac{\chi^2}{2 \cdot n}\left(\frac{1}{g_{Trial}} - \frac{1}{g_{Std}}\right)}$
ΔE^*_{ab}	$t_{\Delta E}$	$\alpha = \frac{\overline{\Delta a^*}}{\overline{\Delta E^*_{ab}}}$ $\beta = \frac{\overline{\Delta b^*}}{\overline{\Delta E^*_{ab}}}$ $\gamma = \frac{\overline{\Delta L^*}}{\Delta E^*_{ab}}$ $g_E = g_{1,1}\alpha^2 + g_{2,2}\beta^2 + g_{3,3}\gamma^2 + 2g_{1,2}\alpha\beta + 2g_{2,3}\beta\gamma + 2g_{1,3}\alpha\gamma$ $\sqrt{\frac{\chi^2}{n \cdot g_E}}$

A1.3 Example Calculations

A1.3.1 Raw data table, $\overline{L^*}_{\mathrm{Std}} = 98.04$, $\overline{a^*}_{\mathrm{Std}} = -0.02$, $\overline{b^*}_{\mathrm{Std}} = 1.78$:

Sample Number	Δa^*	Δb^*	ΔL^*
1	-0.02	0.36	-0.82
2	-0.01	0.37	-0.89
3	-0.04	0.36	-0.73
4	-0.08	0.35	-0.91
5	-0.16	0.36	-0.91
6	0.02	0.31	-0.63
7	-0.01	0.35	-0.64
8	-0.02	0.35	-0.81
9	-0.08	0.45	-0.60
10	-0.01	0.38	-0.68
11	0.00	0.38	-0.61
12	-0.007	0.40	-0.71
13	0.03	0.33	-0.93
14	0.02	0.27	-0.99
15	-0.09	0.41	-0.70
16	-0.01	0.34	-0.87
17	-0.01	0.38	-0.79
18	-0.09	0.39	-0.78
19	-0.05	0.43	-0.87
20	-0.04	0.38	-0.74
Averages	-0.03	0.37	-0.78

A1.3.2

$$\overline{\Delta L^*}_{Trial} = 98.04 + (-0.78) = 97.26 \tag{A1.8}$$

$$\overline{\Delta a^*}_{Trial} = -0.02 + (-0.03) = -0.05 \tag{A1.9}$$

$$\overline{\Delta b^*}_{Trial} = 1.78 + (0.37) = 2.15 \tag{A1.10}$$

$$\overline{\Delta E^*} = \sqrt{(-0.78)^2 + (-0.03)^2 + (0.37)^2} = 0.86 \tag{A1.11}$$

$$\overline{\Delta C^*} = \sqrt{(-0.05)^2 + (2.15)^2} - \sqrt{(-0.02)^2 + (1.78)^2} = 0.37 \tag{A1.12}$$

$$\overline{\Delta h_{ab}} = Arc\ \tan\frac{2.15}{-0.05} - Arc\ \tan\frac{1.78}{-0.02} = 0.95° \tag{A1.13}$$

$$\overline{\Delta H^*} = \sqrt{(0.86)^2 - (-0.78)^2 - (0.37)^2} = 0.10 \qquad \text{(A1.14)}$$

$$\text{var}\,(\Delta a^*) = v_{11} = \frac{1}{20-1}\sum(\Delta a_i^* - \overline{\Delta a^*})^2 = 0.002198 \qquad \text{(A1.15)}$$

$$\text{var}\,(\Delta b^*) = v_{22} = \frac{1}{20-1}\sum(\Delta b_i^* - \overline{\Delta b^*})^2 = 0.001620 \qquad \text{(A1.16)}$$

$$\text{var}\,(\Delta L^*) = v_{33} = \frac{1}{20-1}\sum(\Delta L_i^* - \overline{\Delta L^*})^2 = 0.013689 \qquad \text{(A1.17)}$$

$$\begin{aligned}\text{covar}\,(\Delta a^*, \Delta b^*) &= v_{12}\\ &= v_{21}\\ &= \frac{1}{20-1}\sum(\Delta a_i^* - \overline{\Delta a^*})\cdot(\Delta b_i^* - \overline{\Delta b^*})\\ &= -0.000848 \qquad \text{(A1.18)}\end{aligned}$$

$$\begin{aligned}\text{covar}\,(\Delta a^*, \Delta L^*) &= v_{13}\\ &= v_{31}\\ &= \frac{1}{20-1}\sum(\Delta a_i^* - \overline{\Delta a^*})\cdot(\Delta L_i^* - \overline{\Delta L^*})\\ &= 0.000254 \qquad \text{(A1.19)}\end{aligned}$$

$$\begin{aligned}\text{covar}\,(\Delta b^*, \Delta L^*) &= v_{23}\\ &= v_{32}\\ &= \frac{1}{20-1}\sum(\Delta b_i^* - \overline{\Delta b^*})\cdot(\Delta L_i^* - \overline{\Delta L^*})\\ &= 0.001979 \qquad \text{(A1.20)}\end{aligned}$$

$$V = \begin{bmatrix} 0.002198 & -0.000848 & 0.000254\\ -0.000848 & 0.001620 & 0.001978\\ 0.000254 & 0.001979 & 0.013689 \end{bmatrix}$$

The MINVERSE function of EXCEL applied to the above matrix yields:

$$G = \begin{bmatrix} 622.4854 & 412.8951 & -71.2118\\ 412.8951 & 1023.602 & -155.567\\ -71.2118 & -155.567 & 96.86054 \end{bmatrix}$$

A1.3.3 Using the formulas in Table A1.1 we compute the critical values for each of the color difference components to identify the level of difference that is statistically significant.

A1.3.4 In oral reports to the Inter-Society Color Council and to the Society of Plastics Engineers, it has been reported that statistical acceptance volumes based on the multivariate analysis of production batches using Hotelling's T^2 statistic show very good agreement with the predictions of modern color tolerance equations such as the CMC or CIE94 equations after adjustment for the "commercial factor." Taking a series of readings of a set of material standards using a test instrument and comparing those measures to known values obtained from a standardizing laboratory will provide both a statistical estimate of the confidence volume and an equivalent "color difference" for that probability.

A1.3.4.1 Begin by taking the series of measurements and computing the differences between the measured values and the reference values in terms of ΔL^*, ΔC^* and ΔH^*. Compute the variance-covariance matrix and then the Hotelling's T^2 as follows:

$$T^2 = n(\Delta LCH)'\, S^{-1}\,(\Delta LCH)$$

where the bold letters indicate vectors or matrices. Then Hotelling's T^2 can be related to a noncentral F distribution as:

$$F_{3,n-3} = \frac{(n-3)\,T^2}{3(n-1)}$$

A1.3.4.2 If we let G be the inverse of the variance-covariance matrix (S^{-1}), then the elements of the matrix G will be g_{ij} and the properties of the ellipsoid are given as:

$$\begin{aligned}\Delta E^2 &= g_{11}\,\Delta L^2 + 2g_{12}\Delta C\Delta L + g_{22}\Delta C^2 + 2g_{13}\Delta L\Delta H + 2g_{13}\Delta L\Delta H\\ &\quad + 2g_{23}\Delta C\Delta H + g_{33}\Delta H^2\\ &= (n^{-1})T^2\end{aligned}$$

A1.3.4.3 If CIE94 or CMC is correct, then all cross product terms should be equal to or very near to 0.0 and can be neglected. Then all that needs to be done is to compute the value of cf that scales the ratios of $[g_{11}/(l\,S_L^2)] = [g_{22}/(c\,S_C^2)] = [g_{33}/(h\,S_H^2)] = 1.0$ to determine the correct tolerance size.

TABLE A1.2 Critical Values for 1−α = 95 %, χ^2= 7.81

Color Coordinate	Critical Value	Equations
$\Delta a*$	$t_{\Delta a}$= 0.023	$\sqrt{\frac{\chi^2}{n \cdot g_{1,1}}} = \sqrt{\frac{7.81}{20 \cdot 622.2261}}$
$\Delta b*$	$t_{\Delta b}$= 0.016	$\sqrt{\frac{\chi^2}{n \cdot g_{2,2}}} = \sqrt{\frac{7.81}{20 \cdot 1023.3601}}$
$\Delta L*$	$t_{\Delta L}$= 0.055	$\sqrt{\frac{\chi^2}{n \cdot g_{3,3}}} = \sqrt{\frac{7.81}{20 \cdot 96.8606}}$
ΔH^*_{ab}	$t_{\Delta H}$= 0.022	$f = \frac{1}{2}\Big(\overline{h_{Trial}} - \overline{h_{Standard}} = \frac{1}{2}(90.78° - 91.61°) = 91.20°$ $g_H = g_{1,1} \sin^2 f - 2g_{12} \sin f \cos f + g_{22} \cos^2 f$ $g_H = 622.2261 \cdot 0.9998^2 - 2 \cdot 412.5825 \cdot 0.9998 \cdot (-0.02094) + 1023.3601 \cdot (-0.02094)^2 = 639.7$ $\sqrt{\frac{\chi^2}{n \cdot g_H}} = \sqrt{\frac{7.81}{20 \cdot 639.7}}$
Δh^*_{ab}	$t_{\Delta h}$= 0.21°	$Arc \tan\left(2 \cdot \frac{t_{\Delta H}}{\Delta H^*_{ab}} \cdot \tan\frac{\overline{\Delta h^*_{ab}}}{2}\right) = Arc \tan\left(2 \cdot \frac{0.02241}{0.10} \cdot \tan\frac{0.95}{2}\right)$
ΔC^*_{ab}	$t_{\Delta C}$= 0.016	$g_{Trial} = g_{1,1} \cos^2\overline{h_{Trial}} + 2g_{1,2} \sin\overline{h_{Trial}} \cos\overline{h_{Trial}} + g_{2,2} \sin^2\overline{h_{Trial}}$ $g_{Trial} = 622.2261 \cdot (-0.01361)^2 + 2 \cdot 412.5825 \cdot 0.9999 \cdot (-0.01361) + 1023.3601 \cdot (0.9999)^2 = 1012.0$ $g_{Std} = g_{1,1} \cos^2\overline{h_{Std}} + 2g_{1,2} \sin\overline{h_{Std}} \cos\overline{h_{Std}} + g_{2,2} \sin^2\overline{h_{Std}}$ $g_{Std} = 622.2261 \cdot (-0.0280961)^2 + 2 \cdot 412.5825 \cdot (-0.0280961) \cdot 0.9996 + 1023.3601 \cdot (0.9996)^2 = 999.8$ $\sqrt{\frac{\chi^2}{2 \cdot n}\left(\frac{1}{g_{Trial}} - \frac{1}{g_{Std}}\right)} = \sqrt{\frac{7.81}{2 \cdot 20}\left(\frac{1}{1012.0} - \frac{1}{999.8}\right)}$
ΔE^*_{ab}	$t_{\Delta E}$= 0.0258	$\alpha = \frac{\overline{\Delta a^*}}{\overline{\Delta E^*_{ab}}} = \frac{-0.04}{0.86} = -0.0465$ $\beta = \frac{\overline{\Delta b^*}}{\overline{\Delta E^*_{ab}}} = \frac{0.37}{0.86} = 0.430$ $\gamma = \frac{\overline{\Delta L^*}}{\overline{\Delta E^*_{ab}}} = \frac{-0.77}{0.86} = -0.895$ $g_E = g_{1,1}\alpha^2 + g_{2,2}\beta^2 + g_{3,3}\gamma^2 + 2g_{1,2}\alpha\beta + 2g_{2,3}\beta\gamma + 2g_{1,3}\alpha\gamma$ $g_E = 622261 \cdot (-0.0465)^2 + 1023.3601 \cdot (0.430)^2 + 96.8606 \cdot (-0.895)^2 + 2 \cdot 4125825 \cdot (-0.0465) \cdot (0.430) + 2 \cdot (-155.5824) \cdot (0.430) \cdot (-0.895) + 2 \cdot (-71.1867) \cdot (-0.0465) \cdot (-0.895) = 376.4$ $\sqrt{\frac{\chi^2}{n \cdot g_E}} = \sqrt{\frac{7.81}{20 \cdot 376.4}} = 0.0258$

A2. TEMPERATURE DEPENDENCE OF BCRA CERAMIC TILES

A2.1 The NPL and CERAM have studied the temperature sensitivity of the Ceramic Colour Standards, Series II and have published the following tables of changes in the CIELAB coordinates for a 10°C rise in ambient temperature or the temperature at the surface of the tile.

A2.2 To correct a color measurement to a specific temperature use the following equation:

$$C_{CORRECTED} = C_{MEASURED} + \left(\Delta T_C \times \frac{T_C - T_M}{10} \right) \qquad \text{(A2.1)}$$

Where $C_{CORRECTED}$ is the corrected color coordinates (L^*, a^*, b^*), $C_{MEASURED}$ is the measured color coordinate, ΔT_C is the thermochromic coefficient from either Table A2.1 or Table A2.2, T_M is the temperature at which the measurement was made, and T_C is the temperature to which the color measurement is to be corrected.

Example: Suppose the Yellow tile had been measured at 22.5° C using (8°/t) geometry and resulted in the following CIELAB values (D65/1964) $L^* = 83.50$, $a^* = 1.73$, $b^* = 77.17$. Correct the readings to a standard temperature of 25°C. The calculations are thus:

$$L^* = 83.50 + \left(-0.27 \times \frac{(25 - 22.5)}{10} \right) = 83.43$$

$$a^* = 1.73 + \left(0.70 \times \frac{(25 - 22.5)}{10} \right) = 1.91$$

$$b^* = 77.17 + \left(-0.11 \times \frac{(25 - 22.5)}{10} \right) = 77.14$$

TABLE A2.1 Changes for 10°C Rise in Temperature for Illuminant D65 and the 1964 Standard Observer when Measured with 8°/t (Specular Component Included) Geometry

Tile Color	ΔL^*	Δa^*	Δb^*	ΔE^*
Pale Grey	-0.03	-0.02	0.02	0.04
Mid Grey	-0.03	-0.02	0.04	0.05
Difference Grey	-0.04	0.04	0.03	0.06
Deep Grey	0.00	0.01	0.00	0.01
Deep Pink	-0.10	-0.44	-0.19	0.48
Red	-0.37	-0.71	-0.61	1.01
Orange	-0.45	0.56	-0.66	0.97
Yellow	-0.27	0.70	-0.11	0.76
Green	-0.18	0.66	-0.04	0.68
Difference Green	-0.18	0.69	-0.05	0.72
Cyan	-0.10	0.31	0.01	0.32
Deep Blue	0.00	-0.04	0.05	0.06

TABLE A2.2 Changes for 10°C Rise in Temperature for Illuminant D65 and the 1964 Standard Observer when Measured with 0°/d (Specular Component Excluded) Geometry

Tile Color	ΔL^*	Δa^*	Δb^*	ΔE^*
Pale Grey	-0.03	-0.02	0.03	0.04
Mid Grey	-0.03	-0.03	0.04	0.06
Difference Grey	-0.04	0.04	0.03	0.07
Deep Grey	0.00	0.01	0.00	0.01
Deep Pink	-0.13	-0.48	-0.23	0.55
Red	-0.55	-0.54	-0.83	1.13
Orange	-0.49	0.65	-0.67	1.06
Yellow	-0.29	0.74	0.02	0.79
Green	-0.20	0.75	-0.03	0.77
Difference Green	-0.20	0.77	-0.03	0.80
Cyan	-0.12	0.34	0.00	0.36
Deep Blue	0.01	-0.09	0.08	0.12