Designation: E 1345 – 98

AMERICAN SOCIETY FOR TESTING AND MATERIALS
100 Barr Harbor Dr., West Conshohocken, PA 19428
Reprinted from the Annual Book of ASTM Standards. Copyright ASTM

Standard Practice for Reducing the Effect of Variability of Color Measurement by Use of Multiple Measurements[1]

This standard is issued under the fixed designation E 1345; 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 improvements in the precision and bias of color-measuring instruments have been accompanied by more widespread use of numerical color tolerances based on instrumental measurements. As tighter tolerances are specified, they begin to approach the limits of visual perception. In many cases, the instrument user has found it difficult to prepare and measure specimens with adequate repeatability. This practice provides procedures for reducing variability in the mean results of color measurement by the use of multiple measurements, and it indicates how many measurements are required for a specific reduction.

1. Scope

1.1 Reduction of the variability associated with average color or color-difference measurements of object-color specimens is achieved by statistical analysis of the results of multiple measurements on a single specimen, or by measurement of multiple specimens, whichever is appropriate.

1.2 This practice provides a means for the determination of the number of measurements required to reduce the variability to a predetermined fraction of the relevant color or color-difference tolerances.

1.3 This practice is general in scope rather than specific as to instrument or material.

2. Referenced Documents

2.1 *ASTM Standards:*

D 2244 Test Method for Calculation of Color Differences from Instrumentally Measured Color Coordinates[2]

D 3134 Practice for Establishing Retroreflectance Color and Gloss Tolerances[2]

E 178 Practice for Dealing With Outlying Observations[3]

E 284 Terminology of Appearance[2]

E 308 Practice for Computing the Colors of Objects by Using the CIE System[2]

E 456 Terminology Relating to Quality and Statistics[3]

E 1164 Practice for Obtaining Spectrophotometric Data for Object-Color Evaluation[2]

2.2 *Other Standard:*

SAE J 1545 Recommended Practice for Instrumental Color Difference Measurement for Exterior Finishes, Textiles and Colored Trim[4]

3. Terminology

3.1 Definitions of appearance terms in Terminology E 284 or statistical terms in Terminology E 456 are applicable to this practice.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *box and whisker plot*, *n*—a nonparametric data analysis diagram that illustrates the 25, 50, and 75 % cumulative distribution of values in a data set (the box) and the expected range of values, defined by distance outside the box ends; see *whiskers*, see Fig. 1.

3.2.2 *extreme value*, *n*—a single reading, selected from a series of readings, whose value is farther from the nearer box end than 3.0 times the hinge length.

3.2.2.1 *Discussion*—A box and whiskers plot is normally used to find outliers and extreme values. Such values should be eliminated from a series before calculating the series mean, standard deviation, and confidence intervals.

3.2.3 *hinges*, *n*—the 25 and 75 % cumulative distribution points in a set of readings taken during a measurement.

3.2.3.1 *Discussion*—Hinges represent the values in which 25 % of the readings are less than the lower hinge and 75 % of the readings are less than the upper hinge. See also *hinge length*.

3.2.3.2 *Discussion*—Hinges are sometimes called the lower (Q_1) and upper (Q_1) quartile values.

3.2.4 *hinge length*, *H*, *n*—the range of values between the lower and upper hinges.

[1] This practice is under the jurisdiction of ASTM Committee E-12 on Color and Appearance of Materials and is the direct responsibility of Subcommittee E12.04 on Planning.

Current edition approved June 10, 1998. Published November 1998. Originally published as E1345-90. Last previous edition E1345-90.

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

[3] *Annual Book of ASTM Standards*, Vol 14.02.

[4] Available from the Society of Automotive Engineers, 400 Commonwealth Dr., Warrendale, PA 15096.

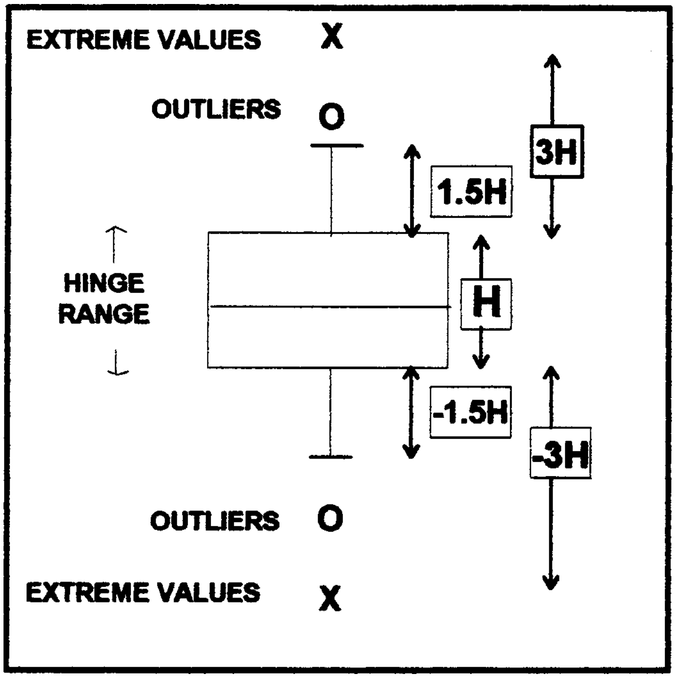

FIG. 1 Schematic Description of a Box and Whisker Plot

3.2.4.1 *Discussion*—The hinge length is sometimes called the box width or the interquartile range Q_3 to Q_1.

3.2.5 *outlier*, *n*—a single reading, selected from a series of readings, whose value is further from the nearer box end then 1.5 times the hinge length; see see 3.2.2.1.

3.2.6 *sampling number, N*, *n*—number of multiple measurements, or number of multiple specimens, required to reduce the variability of color or color-difference measurement to a desired level.

3.2.7 *standard deviation of color or color-difference measurement, s*—standard deviation of the color scale or color-difference-scale value, x_i, being considered:

$$s = [\{\Sigma(x_i - x_{avg})^2\}/(n - 1)]^{0.5} \quad (1)$$

where:

x_{avg} = $(\Sigma\ x_i)/n$, and
n = the number of replicate measurements made.

3.2.8 *standard deviation of instrument*, s_i, *n*—standard deviation of a color-scale or color-difference-scale value due to instrument variability alone:

$$s_i = [\{\Sigma(x_i - x_{avg})^2\}/(n - 1)]^{0.5} \quad (2)$$

3.2.9 *standard error of the estimated mean*, s_e, *n*—standard deviation of color or color-difference measurement divided by the square root of the sampling number:

$$s_e = s/(N^{0.5}) \quad (3)$$

TABLE 1 Appropriate and Inappropriate Color Coordinates for Use in This Practice

Color Coordinates	Appropriate	Inappropriate
CIE *Yxy*		√
CIE *LCH*	√	
CIE *LAB*	√	
CIE LUV	√	
CIE L*u′v′*		√

3.2.10 *standard error goal*, $s_{e,g}$, *n*—level to which the standard error of the estimated mean is to be reduced.

3.2.11 *tolerance*, *n*—the upper tolerance limit minus the lower tolerance limit; the total allowable range of the color-scale or color-difference-scale value considered.

3.2.12 *whiskers*, *n*—lines extending out from the box ends to the largest and smallest observations lying within 1.5 times the hinge length, measured from the box ends.

4. Summary of Practice

4.1 This practice assumes that, for the material under consideration and a specified set of color scales, relevant color or color-difference tolerances have been established (see Practice D 3134).

4.2 For convenience, the numerical example in the Appendix uses CIELAB *LCH* (lightness, chroma, hue) color difference scales ΔL^*, ΔC^*_{ab}, and ΔH^*_{ab} (see Test Method D 2244 and Practice E 308), but this is not meant to be restrictive.

NOTE 1—Some coordinates, such as CIE *x, y, Y*, do not follow the theories of this standard due to excessive colinearity. While it has not been tested, this same colinearity problem may also be observed in 1960 *u, v* and 1976 *u′, v′* coordinates. Table 1 provides a listing of the appropriate and inappropriate color coordinates for use with this practice.

4.3 The successive steps in the procedure are as follows:

4.3.1 Determine the standard deviation of instrument.

4.3.1.1 Screen the measurement data for outliers and extreme values.

4.3.2 Determine the standard deviation of color or color-difference measurement.

4.3.2.1 Screen the measurement data for outliers and extreme values.

4.3.3 Determine the standard error of the estimated mean for a sampling number of one.

4.3.4 Determine the final sampling number that reduces the standard error of the estimated mean to less than the standard error goal for each scale value.

4.3.5 Determine the final standard error goal values.

NOTE 2—When the standard error of the estimated mean for a sampling number of one is larger than a specified fraction of the tolerance or a specified multiple of the standard deviation of instrument for any of the three color-difference-scale values, a sampling number greater than one is required.

4.4 *Screening for and Elimination of Outliers and Extreme Values in Measured Data*:

4.4.1 *Box and whisker test*—This test is best carried out by computer. Many programs for the box and whisker technique are available.[5]

4.4.1.1 Order the readings from lowest to highest value. The reading whose value is half way between the minimum and maximum values is the median. Fig. 1 illustrates the following steps.

4.4.1.2 The reading whose value is just less than 75 % of the other readings is the lower hinge. The readings whose value is just higher than 75 % of the other readings is the upper hinge. The difference between these two is the hinge length *H*.

[5] See for example, Schaefer, R. L. and Anderson, R. B., *The Student Edition of Minitab*, Addison-Wesley, New York, 1989.

4.4.1.3 If the smallest value of any reading is less than the lower hinge value minus 1.5 times the hinge length, it may be considered an outlier. Likewise, if the largest value of any reading is greater than the upper hinge value plus 1.5 times the hinge length, it may be considered an outlier.

4.4.1.4 If the smallest (largest) value of any reading is less (greater) than the lower (upper) hinge value minus (plus) 3.0 times the hinge length, it may be considered an extreme value.

4.4.2 *Practice E 178 Procedure*—The test for outliers in Practice E 178 is constructed from the sample mean X_{avg}, and the standard deviation *s*.

4.4.2.1 Order the readings from lowest to highest value.

4.4.2.2 Calculate the following two statistics, T_1 for the lowest value, and T_n for the highest value in a set of *n* ordered readings as follows:

$$T_l = \frac{(x_{avg} - x_l)}{s} \tag{4}$$

$$T_n = \frac{(x_n - x_{avg})}{S} \tag{5}$$

4.4.2.3 Compare the values of $T_l(T_n)$ to critical values in Table 2. If $T_l(T_n)$ is larger than the critical value for *n* readings at the 1 % level of significance. Reading 1 (*n*) may be considered an outlier.

NOTE 3—Table 2 contains critical values for series of up to 15 readings and for 0.1 and 1 % significance levels. For other significance levels and larger datasets, see Table 1 of Practice E 178.

4.4.2.4 If $T_l(T_n)$ is larger than the critical value for *n* readings at the 1 % level of significance, Readings 1 (*n*) may be considered an extreme value.

4.4.3 If any outliers or extreme values were found, consider carefully whether they should be dropped or retained. Drop those readings not considered to be part of the desired dataset, by whatever consistent criteria are accepted. See 5.3.

4.4.4 Recalculate the mean, standard deviation and confidence limits of the remaining dataset.

5. Significance and Use

5.1 This practice should be used whenever measured color-scale or color-difference-scale values are to be compared to an established tolerance. In this way it can be demonstrated quantitatively that the sampling and measurement procedures are adequate to allow an unambiguous decision as to whether or not the mean results are within tolerance.

TABLE 2 Official Values for *T* (One-Sided Test) for Outliers

Number of Observations *n*	Upper 0.1 % Significance Level	Upper 1.0 % Significance Level
3	1.155	1.155
4	1.499	1.492
5	1.780	1.749
6	2.011	1.944
7	2.201	2.097
8	2.358	2.221
9	2.492	2.323
10	2.606	2.410
11	2.705	2.485
12	2.791	2.550
13	2.867	2.607
14	2.935	2.659
15	2.997	2.705

5.2 This practice is based on portions of SAE Practice J 1545, as it applies to painted or plastic automotive parts. It is generally applicable to object colors in various materials. Textured materials, such as textiles, may require special consideration (see SAE Practice J 1545 and STP 15D Manual on Presentation of Data and Control Chart Analysis[6]).

5.3 While Practice E 178 deals with outliers, it does not include definitions relating to the box and whisker technique. The definition of an outlier is operational and a little vague because there is still considerable disagreement about what constitutes an outlier. In any normally distributed population, there will be members that range from minus to plus infinity. Theoretically, one should include any member of the population in any sample based on estimates of the population parameters. Practically, including a member that is found far from the mean within a small sample, most members of which are found near the mean, will introduce a systematic bias into the estimate of the population parameters (mean, standard deviation, standard error). Such a bias is in direct contrast with the goal of this practice, namely, to reduce the effects of variability of measurement. For the purposes of this practice, no distinction is made between errors of sampling and members of the tails of the distribution. Practice E 178 has several methods and significance tables to attempt to differentiate between these two types of extreme values.

6. Procedure

6.1 Determine the standard deviation of instrument, s_i, by carrying out the appropriate color measurement at least 10 times ($n = 10$) when using a stable product standard as the specimen, without removing or disturbing the specimen between measurements. Calculate s_i by the use of Eq 2. This determination should be carried out for each color scale used and for each product with a new color; however, s_i is unlikely to change appreciably over relatively extended periods.

6.1.1 Screen the measurement data for outliers and extreme values following 4.4.1-4.4.4.

6.2 Select maximum allowable values of the standard error of the estimated mean, as a fraction of the tolerance and as a multiple of the standard deviation of instrument. In the absence of specified values of these quantities, use those recommended in SAE Practice J 1545: 0.1 times the tolerance and $2s_i$. These values are used in the Appendix.

NOTE 4—This practice assumes that all measurements are subject to the central limit theorem of mathematical statistics, so that as the number of replicate or repeat measurements becomes large, the distribution of values is described by the standard normal distribution. It has been shown,[7,8] however, that averages of large numbers of measurements of a verification standard on a properly maintained spectrophotometer are not approximated by the standard normal distribution. As a result, tests anchored to s_i may exhibit a significance or a power dependence different from that which is expected.

[6] Available from ASTM Headquarters 100 Barr Harbor Drive, West Conshohocken, PA 19428.

[7] Marcus, R. T., and Billmeyer, F. W., Jr., "Statistical Study of Color-Measuring Instruments," *Applied Optics*, Vol 13, 1974, pp. 1519–1530.

[8] Billmeyer, F. W., Jr., and Alessi, P. J., "Assessment of Color-Measuring Instruments,'' *Color Research and Application*, Vol 6, 1981, pp. 195–202.

6.3 Determine the standard deviation of color or color-difference measurement, s, by making the appropriate measurement at least 10 times ($n = 10$), as follows:

6.3.1 To assess the variability within a single specimen, measure the same specimen at ten or more randomly selected different areas of the specimen.

6.3.1.1 Screen the measurement data for outliers and extreme values following 4.4.1-4.4.4.

6.3.2 To assess the variability among specimens, measure at least ten replicate specimens.

6.3.2.1 Screen the measurement data for outliers and extreme values following 4.4.1-4.4.4.

6.4 Determine the standard error of the estimated mean, s_e, for a sampling number of one, using Eq 3. Note that for $N = 1$, $s_e = s$. Use the larger of the values of s determined in 6.3.1 or 6.3.2.

6.5 Compare the value of s_e to 0.1 times the tolerance and to $2s_i$ for each of the three color or color-difference scales used. When in any of the three cases s_e exceeds $2s_i$ or 0.1 times the tolerance, multiple measurements are required ($N > 1$). Whether these should be multiple measurements of a single specimen or measurements of multiple (replicate) specimens is determined by whether the value of s from 6.3.1 or 6.3.2 is greater.

6.6 Determine the value of the standard error goal, $s_{e,g}$, as the greater of $2s_i$ or 0.1 times the tolerance, for each color or color-difference scale used.

6.7 Calculate the sampling number required to meet the criteria of $s_{e,g}$ as follows:

6.7.1 For each color or color-difference scale, calculated N by the following rearrangement of Eq 3:

$$N = (s/s_{e,g}) \quad (6)$$

6.7.2 Round any fractional values of N to the next larger whole number.

6.7.3 Select the largest of the rounded values of N as the final sampling number.

6.8 Using the final value of N from 6.7.3, calculate the final standard error goal for each color scale by use of Eq 3.

7. Report

7.1 Report the final sampling number from 6.7.3 and the final standard error goal for each color scale from 6.8, in addition to the quantities required in the report of the test method used.

8. Precision and Bias

8.1 *Precision and Bias of Final Sampling Number, N*—Since N has been rounded up to the next larger whole number, its precision is ±1 unit and its maximum bias is + 1 unit.

8.2 *Precision and Bias of Final Standard Error Goals*, $s_{e,g}$:

8.2.1 The calculations of this practice can affect the precision of $s_{e,g}$ due to roundoff. To minimize this error, one more significant figure should be carried in the calculations than is required by the precision and bias statement of the test method used.

8.2.2 The calculations of this practice should introduce no bias into $s_{e,g}$.

8.2.3 To the quantities of 8.2.1 should be added any contribution to precision or bias resulting from the test method used.

9. Keywords

9.1 color; color difference; color measurement; color tolerances

APPENDIX

(Nonmandatory Information)

X1. CALCULATION OF THE FINAL SAMPLING NUMBER AND THE FINAL STANDARD ERROR OF THE ESTIMATED MEAN

TABLE X1.1

Section	Quantity	Color-Difference Scale		
		ΔL^*	ΔC^*_{ab}	ΔH^*_{ab}
6.1	Instrument standard deviation, s_i	0.1	0.1	0.1
6.2	Least significant scale-value interval, $2s_i$	0.2	0.2	0.2
	Upper tolerance limit	+ 2.0	+ 1.0	+ 0.5
	Lower tolerance limit	−2.0	−1.0	−0.5
	Tolerance	4.0	2.0	1.0
	0.1 times the tolerance	0.4	0.2	0.1
6.3	Standard deviation, s	0.45	0.35	0.15
6.4	Standard error of estimated mean, s_e, for $N = 1$	0.45	0.35	0.15
6.6	Standard error goal, $s_{e,g}$	0.4	0.2	0.2
6.7.1	Sampling number, N	1.27	3.06	0.56
6.7.2	Rounded sampling number	2	4	1
6.7.3	Final sampling number	4	4	4
6.8	Final standard error goal	0.23	0.18	0.08