Designation: E2056 – 04 (Reapproved 2016)

Standard Practice for Qualifying Spectrometers and Spectrophotometers for Use in Multivariate Analyses, Calibrated Using Surrogate Mixtures[1]

This standard is issued under the fixed designation E2056; 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.

1. Scope

1.1 This practice relates to the multivariate calibration of spectrometers and spectrophotometers used in determining the physical and chemical characteristics of materials. A detailed description of general multivariate analysis is given in Practices E1655. This standard refers only to those instances where surrogate mixtures can be used to establish a suitable calibration matrix. This practice specifies calibration and qualification data set requirements for interlaboratory studies (ILSs), that is, round robins, of standard test methods employing surrogate calibration techniques that do not conform exactly to Practices E1655.

NOTE 1—For some multivariate spectroscopic analyses, interferences and matrix effects are sufficiently small that it is possible to calibrate using mixtures that contain substantially fewer chemical components than the samples that will ultimately be analyzed. While these surrogate methods generally make use of the multivariate mathematics described in Practices E1655, they do not conform to procedures described therein, specifically with respect to the handling of outliers.

1.2 This practice specifies how the ILS data is treated to establish spectrometer/spectrophotometer performance qualification requirements to be incorporated into standard test methods.

NOTE 2—Spectrometer/spectrophotometer qualification procedures are intended to allow the user to determine if the performance of a specific spectrometer/spectrophotometer is adequate to conduct the analysis so as to obtain results consistent with the published test method precision.

1.2.1 The spectroscopies used in the surrogate test methods would include but not be limited to mid- and near-infrared, ultraviolet/visible, fluorescence and Raman spectroscopies.

1.2.2 The surrogate calibrations covered in this practice are: multilinear regression (MLR), principal components regression (PCR) or partial least squares (PLS) mathematics. These calibration procedures are described in detail in Practices E1655.

1.3 For surrogate test methods, this practice recommends limitations that should be placed on calibration options that are allowed in the test method. Specifically, this practice recommends that the test method developer demonstrate that all calibrations that are allowed in the test method produce statistically indistinguishable results.

1.4 For surrogate test methods that reference spectrometer/spectrophotometer performance practices, such as Practices E275, E925, E932, E958, E1421, E1683, or E1944; Test Methods E387, E388, or E579; or Guide E1866, this practice recommends that instrument performance data be collected as part of the ILS to establish the relationship between spectrometer/spectrophotometer performance and test method precision.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D6277 Test Method for Determination of Benzene in Spark-Ignition Engine Fuels Using Mid Infrared Spectroscopy

D6300 Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products and Lubricants

E131 Terminology Relating to Molecular Spectroscopy

E275 Practice for Describing and Measuring Performance of Ultraviolet and Visible Spectrophotometers

E387 Test Method for Estimating Stray Radiant Power Ratio of Dispersive Spectrophotometers by the Opaque Filter Method

E388 Test Method for Wavelength Accuracy and Spectral Bandwidth of Fluorescence Spectrometers

E579 Test Method for Limit of Detection of Fluorescence of Quinine Sulfate in Solution

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

E925 Practice for Monitoring the Calibration of Ultraviolet-Visible Spectrophotometers whose Spectral Bandwidth

[1] This practice is under the jurisdiction of ASTM Committee E13 on Molecular Spectroscopy and Separation Science and is the direct responsibility of Subcommittee E13.11 on Multivariate Analysis.

Current edition approved April 1, 2016. Published May 2016. Originally approved in 1999. Last previous edition approved in 2010 as E2056 – 04(2010). DOI: 10.1520/E2056-04R16.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

does not Exceed 2 nm
E932 Practice for Describing and Measuring Performance of Dispersive Infrared Spectrometers
E958 Practice for Estimation of the Spectral Bandwidth of Ultraviolet-Visible Spectrophotometers
E1421 Practice for Describing and Measuring Performance of Fourier Transform Mid-Infrared (FT-MIR) Spectrometers: Level Zero and Level One Tests
E1655 Practices for Infrared Multivariate Quantitative Analysis
E1683 Practice for Testing the Performance of Scanning Raman Spectrometers
E1866 Guide for Establishing Spectrophotometer Performance Tests
E1944 Practice for Describing and Measuring Performance of Laboratory Fourier Transform Near-Infrared (FT-NIR) Spectrometers: Level Zero and Level One Tests

3. Terminology

3.1 *Definitions:*

3.1.1 For definitions of terms and symbols relating to infrared, ultraviolet/visible and Raman spectroscopy, refer to Terminology E131.

3.1.2 For definitions of terms and symbols relating to multivariate analysis, refer to Practices E1655.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *spectrometer/spectrophotometer qualification, n*—the procedures by which a user demonstrates that the performance of a specific spectrometer/spectrophotometer is adequate to conduct a multivariate analysis so as to obtain precision consistent with that specified in the test method.

3.2.2 *surrogate calibration, n*—a multivariate calibration that is developed using a calibration set which consists of mixtures with pre-specified and reproducible compositions that contain substantially fewer chemical components than the samples that will ultimately be analyzed.

3.2.3 *surrogate test method, n*—a standard test method that is based on a surrogate calibration.

4. Summary of Practice

4.1 A surrogate test method must specify the composition of two sets of samples. One set is used to calibrate the spectrometers/spectrophotometers. The second set of samples is used to qualify the spectrometer/spectrophotometer to perform the analysis. The compositions of both sets are expressed in terms of weight or volume fraction depending on whether the samples are prepared gravimetrically or volumetrically. The compositions of both sets should be specified in the surrogate test method. If the surrogate test method is being used to estimate a physical property, then the test method should indicate what value of the property is to be assigned to each of the calibration and qualification samples.

4.2 The surrogate test method should specify the minimum spectrometer/spectrophotometer requirements for instruments that can be used to perform the test method.

4.3 The spectrometer/spectrophotometer test method should specify the exact conditions that are to be used to collect and, where appropriate, to calculate the spectral data used in the calibration and analysis.

4.4 The test method should specify the exact mathematics that are to be used to develop the multivariate calibration. Allowable spectral preprocessing methods should be defined. The specific mathematics (MLR, PCR or PLS) should be specified, and the acceptable range for the numbers of variables should be given.

4.5 When the ILS is conducted to establish the precision of the surrogate test method, the calibration data for all of the participating laboratories should be collected and used to calculate a pooled standard error of calibration for the test method. The pooled standard error of calibration and its associated degrees of freedom should be reported in the test method.

4.5.1 When a user is calibrating a spectrometer/spectrophotometer, the standard error of calibration is calculated and compared to the pooled standard error of calibration from the ILS to determine if the performance of the calibrated spectrometer/spectrophotometer is adequate to produce analyses of the precision specified in the test method.

4.5.2 If a user is purchasing a precalibrated spectrometer/spectrophotometer, the instrument vendor should supply the standard error of calibration and its statistical comparison to the pooled standard error of calibration.

4.6 During the ILS, each participating laboratory analyzes a set of qualification samples and reports both the compositions of the qualification set and the estimates made using the multivariate analysis. A pooled error of qualification is calculated and reported as part of the test method along with its corresponding degrees of freedom.

4.6.1 Before a user may use the spectrometer/spectrophotometer, it must be qualified to perform the surrogate test method. The qualification set is analyzed, and a standard error of qualification is calculated. The standard error of qualification is statistically compared with the pooled standard error of qualification to determine if the performance of the calibrated spectrometer/spectrophotometer is adequate to produce analyses of the precision specified in the test method.

4.6.2 Spectrometer/spectrophotometer qualification is required regardless of whether the calibration is performed by the vendor or the user.

4.6.3 Spectrometer/spectrophotometer qualification should be repeated after major maintenance has been performed on the spectrometer/spectrophotometer so as to determine whether recalibration is required.

5. Significance and Use

5.1 This practice should be used by the developer of standard test methods that employ surrogate calibrations.

5.1.1 This practice assists the test method developer in setting and documenting requirements for the spectrometer/spectrophotometers that can perform the test method.

5.1.2 This practice assists the test method developer in setting and documenting spectral data collection and computation parameters for the test method.

5.1.3 This practice assists the test method developer in selecting among possible multivariate analysis procedures that could be used to establish the surrogate calibration. The practice describes statistical tests that should be performed to ensure that all multivariate analysis procedures that are allowed within the scope of the test method produce statistically indistinguishable results.

5.1.4 This practice describes statistical calculations that the test method developer should perform on the calibration and qualification data that should be collected as part of the ILS that establishes the test method precision. These calculations establish the level of performance that spectrometers/spectrophotometers must meet in order to perform the test method.

5.2 This practice describes how the person who calibrates a spectrometer/spectrophotometer can test the performance of said spectrometer/spectrophotometer to determine if the performance is adequate to conduct the test method.

5.3 This practice describes how the user of a spectrometer/spectrophotometer can qualify the spectrometer/spectrophotometer to conduct the test method.

6. Surrogate Calibrations

6.1 Practices E1655 assumes that the calibration set used to develop a multivariate model contains samples of the same type as those that are to eventually be analyzed using the model. Practices E1655 requires use of outlier statistics to ensure that samples being analyzed are sufficiently similar to the calibration samples to produce meaningful results. For some spectroscopic analyses, however, it is possible to calibrate using gravimetrically or volumetrically prepared mixtures that contain significantly fewer components than the samples that will ultimately be analyzed. For these surrogate test methods, the outlier statistics described in Practices E1655 are not appropriate since all samples are expected to be outliers relative to the simplified calibrations. Thus, surrogate test methods cannot fulfill the requirements of Practices E1655. While surrogate test methods may make use of the mathematics described in Practices E1655, they should not claim to follow the procedures described in that practice.

6.1.1 In developing surrogate test methods, it is necessary to thoroughly understand and account for potential spectral interferences. Typically, the spectral range used in surrogate calibrations will be limited so as to minimize interferences. For those interferences that cannot be eliminated through limiting the spectral range, representative components that mimic the interference should be included in the calibration mixtures.

6.1.2 Test Method D6277 provides an example of a surrogate test method. The FT-MIR analysis of benzene in gasoline is calibrated using mixtures of benzene, isooctane, toluene and xylenes and PLS mathematics. The calibration mixtures contain far fewer components than gasoline, but the spectral range used in the analysis is limited to a narrow range about a relatively interference-free benzene peak. Toluene and xylenes are used in the calibration mixtures to adequately mimic the interferences that are present in gasolines.

6.2 *Calibration Sets:*

6.2.1 The sets of surrogate samples that are used to calibrate the spectrometers/spectrophotometers should satisfy the requirements of Practices E1655. If k is the number of variables (MLR wavelengths or frequencies, PCR principal components or PLS latent variables) used in the model, then the minimum number of calibration samples should be the greater of 24 or $6k$. If the calibration set is derived from an experimental design, and if the spectra have been shown to be linear functions of the component concentrations, then fewer calibration samples can be used, but in all cases the minimum number of calibration samples should be the greater of 24 or $4k$. The experimental design must independently vary all components over the desired analysis range.

6.2.2 When calibrating for a single component, the calibration set should uniformly span the range over which the analysis of that component is to be conducted. Additional components that are present in the calibration set to simulate interferences should be independently and uniformly varied over a range at least as large as is likely to be encountered during actual application of the test method.

6.2.3 When calibrating for a property that depends on more than one chemical component, the calibration set should uniformly span the range over which the property analysis is to be conducted, and all components that contribute to the property should be varied independently.

6.2.4 The test method should specify the compositions of the calibration samples, including components and target concentrations. The purity of materials to be used in preparing the calibration samples should also be specified in the test method.

6.3 *Qualification Sets:*

6.3.1 The sets of surrogate samples that are used to qualify the spectrometers/spectrophotometers should satisfy the validation requirements of Practices E1655. If k is the number of variables (MLR wavelengths or frequencies, PCR principal components or PLS latent variables) used in the model, then the minimum number of qualification samples should be the greater of 20 or $5k$. If the qualification set is derived from an experimental design, and if the spectra have been shown to be linear functions of the component concentrations, then fewer qualification samples can be used, but in all cases the minimum number of qualification samples should be the greater of 20 or $3k$. The experimental design must independently vary all components over the entire calibration range.

6.3.2 The compositions of the qualification samples should span the same ranges as did the calibration samples.

6.3.3 The test method should specify the compositions of the qualification samples, including components and target concentrations. The purity of materials to be used in preparing the qualification samples should also be specified in the test method.

6.4 *Precision of Surrogate Calibration Test Methods:*

6.4.1 An ILS determines the precision of a surrogate test method. The interlaboratory study must conform to the requirements of Practice E691, and to any other relevant practices. For example, a test method applicable to petroleum products should conform to Practice D6300.

6.4.2 The standard error of calibration ($SEC_{surrogate}$) and the standard error of qualification ($SEQ_{surrogate}$) for a surrogate test method cannot be used reliably to infer the precision that can be expected for the analysis of actual samples. However, $SEC_{surrogate}$ and $SEQ_{surrogate}$ are representative of the necessary spectrometer/spectrophotometer performance that must be achieved in order to obtain precision comparable to that established by the ILS.

7. Requirements for Test Methods Using Surrogate Calibrations

7.1 *Surrogate Calibrations of Individual Spectrometers/Spectrophotometers:*

7.1.1 The multivariate spectroscopic analysis is calibrated using a set of surrogate mixtures. These mixtures are prepared volumetrically or gravimetrically to compositions defined by the test method. Spectra of the mixtures are collected under conditions defined by the test method. The spectral data is pretreated as prescribed in the test method, and a multivariate calibration model is developed as prescribed in the test method.

7.1.1.1 The y values that are used in the development of the model can be the concentrations of individual components in the surrogate mixtures, or the sum of component concentrations depending on the application.

7.1.1.2 For some applications, the y values that are used in the calibration may be property values that can be calculated from the compositions of the mixtures.

NOTE 3—For some surrogate calibrations, it may be possible to establish a correlation equation that relates the surrogate analyses to results from another analytical test method. It is recommended that multiplicative or additive factors determined from such a correlation not be incorporated into the y values of the surrogate calibration. Instead, the y values should consist of the actual component concentrations, the surrogate test method results should be reported in terms of these concentrations, and the test method should contain a separate section that compares the two test methods and gives the correlation equation.

7.1.2 A standard error of calibration for the surrogate calibration is calculated as:

$$SEC_{surrogate} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{DOF}} \quad (1)$$

where:

DOF = the number of degrees of freedom for the calibration and is $n–k–1$ if the model is mean centered, and $n–k$ otherwise,
n = the number of surrogate mixtures used in the calibration,
k = the number of variables (MLR wavelengths or frequencies, PCR principal components, or PLS latent variables) used in the model,
y_i = the component concentration for the i^{th} calibration sample, and
$\hat{y}_i$ = the estimate of the concentration of the i^{th} calibration sample.

7.2 *Pooled Standard Error of Calibration:*

7.2.1 During the interlaboratory study that establishes the precision of the surrogate test method, each of the m participating laboratories should report a complete set of calibration results consisting of the following:

7.2.1.1 The component concentration or property for the i^{th} calibration sample from the j^{th} laboratory, denoted as y_{ij},

7.2.1.2 The estimate of the concentration of the i^{th} calibration sample from the j^{th} laboratory obtained using the multivariate model to analyze the calibration spectrum, denoted as $\hat{y}_{ij}$,

7.2.1.3 The number of calibration samples for the j^{th} laboratory, denoted as n_j, and

7.2.1.4 The number of variables used in the multivariate model for the j^{th} laboratory, denoted as k_j.

7.2.2 The pooled standard error of calibration is calculated as:

$$PSEC_{surrogate} = \sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n} (\hat{y}_{ij} - y_{ij})^2}{\sum_{j=1}^{m} n_j - k_j - \delta_j}} \quad (2)$$

The sum with index j is over the m laboratories, and δ_j is 1 for labs that use a mean-centered calibration and 0 for labs whose calibration is not mean-centered.

7.2.3 The degrees of freedom for the pooled standard error of calibration, $DOF(PSEC_{surrogate})$, is calculated as:

$$DOF(PSEC_{surrogate}) = \sum_{j=1}^{m} n_j - k_j - \delta_j \quad (3)$$

7.2.4 The surrogate test method should document both $PSEC_{surrogate}$ and $DOF(PSEC_{surrogate})$.

7.3 *Determining Adequacy of Spectrometer/Spectrophotometer Calibrations*—The surrogate test method should indicate that, when a spectrometer/spectrophotometer is calibrated either by an end user or a vendor, the adequacy of the calibration is tested by comparing $SEC_{surrogate}$ with $PSEC_{surrogate}$. The comparison is done using an F-test. The $F_{calibration}$ value is calculated as:

$$F_{calibration} = \frac{SEC^2_{surrogate}}{PSEC^2_{surrogate}} \quad (4)$$

The calculated $F_{calibration}$ value is compared to the critical F value from Table 1 for DOF (see 6.1.2) degrees of freedom in the numerator and $DOF(PSEC_{surrogate})$ (see 6.2.3) in the denominator.

7.3.1 If the calculated $F_{calibration}$ value is less than or equal to the critical F value, then the calibration of the spectrometer/spectrophotometer is comparable to or better than those that participated in the ILS, and the user may continue with the qualification of the spectrometer/spectrophotometer.

7.3.2 If the calculated $F_{calibration}$ value is greater than the critical F value, then the calibration is poorer than those that participated in the ILS. The cause of the poorer performance should be identified and corrected, and the spectrometer/spectrophotometer should be recalibrated.

NOTE 4—The F-test in 7.3.1 is a one-sided test conducted at the 95 % level. The test is one-sided since it is only necessary to show that the variance for the current calibration ($SEC_{surrogate}^2$) is not worse than that for the calibrations used in the interlaboratory study ($PSEC_{surrogate}^2$). If $SEC_{surrogate}^2$ and $PSEC_{surrogate}^2$ come from the same population, then

TABLE 1 95 Percentiles of the *F* Statistic (One-Sided Test)

Denominator, Degrees of Freedom	Numerator 7	8	9	10	12	14	16	18	20	25	30	40	50	100
7	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44	3.40	3.38	3.34	3.32	3.27
8	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15	3.11	3.08	3.04	3.02	2.97
9	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94	2.89	2.86	2.83	2.80	2.76
10	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77	2.73	2.70	2.66	2.64	2.59
11	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65	2.60	2.57	2.53	2.51	2.46
12	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54	2.50	2.47	2.43	2.40	2.35
13	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46	2.41	2.38	2.34	2.31	2.26
14	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39	2.34	2.31	2.27	2.24	2.19
15	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33	2.28	2.25	2.20	2.18	2.12
16	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28	2.23	2.19	2.15	2.12	2.07
17	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23	2.18	2.15	2.10	2.08	2.02
18	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19	2.14	2.11	2.06	2.04	1.98
19	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16	2.11	2.07	2.03	2.00	1.94
20	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12	2.07	2.04	1.99	1.97	1.91
25	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01	1.96	1.92	1.87	1.84	1.78
30	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93	1.88	1.84	1.79	1.76	1.70
35	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88	1.82	1.79	1.74	1.70	1.63
40	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84	1.78	1.74	1.69	1.66	1.59
45	2.22	2.15	2.10	2.05	1.97	1.92	1.87	1.84	1.81	1.75	1.71	1.66	1.63	1.55
50	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78	1.73	1.69	1.63	1.60	1.52
60	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75	1.69	1.65	1.59	1.56	1.48
70	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72	1.66	1.62	1.57	1.53	1.45
80	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70	1.64	1.60	1.54	1.51	1.43
90	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69	1.63	1.59	1.53	1.49	1.41

there is only a 5 % chance that the $F_{calibration}$ will be greater than the value in Table 1.

7.4 *Standard Error of Qualification for Individual Spectrometers/Spectrophotometers:*

7.4.1 Before a spectrophotometer can be used to analyze actual samples, it must be qualified. A qualification set of surrogate mixtures are prepared volumetrically or gravimetrically to compositions defined by the test method. Spectra of the qualification mixtures are collected under conditions defined by the test method. The spectral data is pretreated as prescribed in the test method, and analyzed using the multivariate calibration model as described in the test method.

7.4.2 A standard error of qualification is calculated as:

$$SEQ_{surrogate} = \sqrt{\frac{\sum_{i=1}^{q} (\hat{y}_i - y_i)^2}{q}} \quad (5)$$

where:

q = the number of surrogate qualification mixtures,
y_i = the component concentration for the i^{th} qualification sample, and
$\hat{y}_i$ = the estimate of the concentration of the i^{th} qualification sample.

7.5 *Pooled Standard Error of Qualification*—During the interlaboratory study that establishes the precision of the surrogate test method, each of the m participating laboratories should report a complete set of qualification results consisting of the following:

7.5.1 The component concentration or property for the i^{th} qualification sample from the j^{th} laboratory, denoted as y_{ij},

7.5.2 The estimate of the concentration of the i^{th} qualification sample from the j^{th} laboratory obtained using the multivariate model to analyze the qualification spectrum, denoted as $\hat{y}_{ij}$, and

7.5.3 The number of qualification samples analyzed by the j^{th} laboratory, denoted as q_j.

7.5.4 The pooled standard error of qualification is calculated as:

$$PSEQ_{surrogate} = \sqrt{\frac{\sum_{j=1}^{m} \sum_{i=1}^{q_j} (\hat{y}_{ij} - y_{ij})^2}{\sum_{j=1}^{m} q_j}} \quad (6)$$

7.5.5 The degrees of freedom for the pooled standard error of calibration, $DOF(PSEC_{surrogate})$, is calculated as:

$$DOF(PSEQ_{surrogate}) = \sum_{j=1}^{m} q_j \quad (7)$$

7.5.6 The surrogate test method should document both $PSEQ_{surrogate}$ and $DOF(PSEQ_{surrogate})$.

7.6 *Qualification of an Individual Spectrometer/Spectrophotometer*—The surrogate test method should indicate that, when a spectrometer/spectrophotometer is qualified by an end user, the performance of the calibrated spectrometer/spectrophotometer is tested by comparing $SEQ_{surrogate}$ with $PSEQ_{surrogate}$. The comparison is done via an F-test. The $F_{qualification}$ value is calculated as:

$$F_{qualification} = \frac{SEQ^2_{surrogate}}{PSEQ^2_{surrogate}} \quad (8)$$

The calculated $F_{qualification}$ value is compared to the critical F value from Table 1 for q degrees of freedom in the numerator, and $DOF(SEQ_{surrogate})$ degrees of freedom in the denominator.

7.6.1 If the calculated $F_{qualification}$ value is less than or equal to the critical F value, then the qualification data for the spectrometer/spectrophotometer is comparable to or better than that obtained by laboratories that participated in the ILS. The

user may use the spectrometer/spectrophotometer to conduct analyses in accordance with the surrogate test method.

7.6.2 If the calculated $F_{qualification}$ value is greater than the critical F value, then the qualification data is poorer than that for laboratories that participated in the ILS. The cause of the poorer performance should be identified and corrected, and the spectrometer/spectrophotometer should be recalibrated.

NOTE 5—The F-test in 7.6.1 is also a one-sided test conducted at the 95 % level. The test is one-sided since it is only necessary to show that the variance for the current instrument qualification ($SEQ_{surrogate}^{2}$) is not worse than that for the qualification of instruments used in the interlaboratory study ($PSEQ_{surrogate}^{2}$). If $SEQ_{surrogate}^{2}$ and $PSEQ_{surrogate}^{2}$ come from the same population, then there is only a 5 % chance that the $F_{qualification}$ will be greater than the value in Table 1.

8. Spectrometer/Spectrophotometer Requirements and Performance Tests

8.1 The surrogate test method should contain an apparatus section that details requirements for the spectrometer/spectrophotometers that can be used to conduct the test method analysis. The surrogate test method should reference instrument performance standards by which acceptable spectrometer/spectrophotometer performance is determined. Where possible, spectrometer/spectrophotometer performance data should be collected during the ILS that establishes the precision of the surrogate test method. This data is used to demonstrate that the participating instruments meet the proposed performance requirements.

8.2 *FT-IR Spectrophotometers:*

8.2.1 The surrogate test method should indicate which types of beamsplitters, sources and detectors are permitted for use with the test method.

8.2.2 The surrogate test method should specify a spectral range over which the spectrophotometer is to operate. Typically, the spectral range will be considered to be the frequency range over which the single beam energy spectrum exceeds 10 % of its maximum value.

8.2.3 The surrogate test method should specify a spectral resolution in wavenumbers that is to be used for data collection.

8.2.4 The surrogate test method should reference Practices E1421 or E1944 depending on whether the surrogate test method is a mid- or near-infrared test method.

8.2.4.1 The surrogate test method should specify a maximum allowable noise level measured in some specific frequency range. The frequency ranges used need not correspond to those in Practices E1421 or E1944 if the ranges suggested therein do not correspond to those used in the surrogate calibration model. The noise measurement is typically done on a 100 % line spectrum obtained by ratioing two successive single beam background spectra. Root mean square noise is typically measured over some frequency interval after subtraction of an average transmittance signal. If a maximum allowable noise level is specified, then noise level tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.2.4.2 The surrogate test method should specify a maximum allowable nonphysical energy as a percentage of the single beam maximum energy. The nonphysical energy measurements in Practices E1421 and E1944 are sensitive tests of spectrophotometer linearity. If a maximum allowable nonphysical energy level is specified, then nonphysical energy level tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.2.4.3 If the spectrophotometer must be purged to perform the analysis, then a maximum allowable water vapor level or carbon dioxide level, or both, should be specified. If a maximum allowable water vapor level or carbon dioxide level, or both, is specified, then water vapor tests or carbon dioxide tests, or both, should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.3 *Dispersive Infrared and Ultraviolet-Visible Spectrophotometers:*

8.3.1 The surrogate test method should define source or detector requirements for performing the analysis.

8.3.2 A surrogate test method that employs dispersive infrared spectrophotometers should reference Practices E275 or E932 depending on whether the test method is a near- or mid-infrared test method. A surrogate test method which employs UV-visible spectrophotometers will typically reference Practices E275, E925, or E958, or Test Method E387, or a combination thereof.

8.3.2.1 The surrogate test method should specify a required wavelength or frequency accuracy or precision, or both. The test method should also specify the standard reference material to be used for checking the wavelength or frequency, whether the check is performed on transmittance or absorbance spectra, and the peak finding algorithm to be used to determine the peak positions. If a required wavelength or frequency accuracy or precision, or both, is specified, then wavelength or frequency accuracy tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.3.2.2 The surrogate test method should specify a maximum allowable spectral slit width or spectral bandwidth. A test method for testing spectral slit width or bandwidth should be specified, and these tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.3.2.3 The surrogate test method should specify a photometric precision that is required to perform the analysis. The precision should typically be specified as the standard deviation observed at a specific signal level for a specified number of replicate measurements. If the surrogate test method specifies a required photometric precision, then photometric precision tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.3.2.4 The surrogate test method should specify a required linearity of absorbance or a maximum allowable stray radiant power, or both. The test method should reference appropriate practices for how these performance parameters are measured, and such measurements should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.4 *Fluorescence Spectrophotometers:*

8.4.1 The surrogate test method should define source or detector requirements for performing the analysis.

8.4.2 A surrogate test method that employs fluorescence spectrophotometers should typically reference Test Methods E388 or E579, or both.

8.4.2.1 The surrogate test method should specify a required wavelength accuracy or precision, or both. The test method should also specify the standard reference material to be used for checking the wavelength or frequency, whether the check is performed on transmittance or absorbance spectra, and the peak finding algorithm to be used to determine the peak positions. If a required wavelength accuracy or precision, or both, is specified, then wavelength accuracy tests or precision tests, or both, should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.4.2.2 The surrogate test method should specify a maximum allowable spectral slit width or spectral bandwidth. A test method for testing spectral slit width or bandwidth should be specified, and these tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.4.2.3 The surrogate test method should specify a minimum sensitivity required to perform the analysis. The test method for testing the sensitivity should be specified. If a minimum sensitivity is specified, then tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.5 *Raman Spectrometers:*

8.5.1 The surrogate test method should define source or detector requirements for performing the analysis.

8.5.1.1 The surrogate test method should specify a required frequency accuracy or precision, or both. The test method should also specify the standard reference material to be used for checking the wavelength or frequency, and the peak finding algorithm to be used to determine the peak positions. If a required frequency accuracy or precision, or both, is specified, then frequency accuracy tests or precision tests, or both, should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.5.1.2 The surrogate test method should specify a maximum allowable spectral slit width or spectral bandwidth. A test method for testing spectral slit width or bandwidth should be specified, and these tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

8.5.1.3 The surrogate test method should specify a maximum allowable dark signal level. Dark signal level tests should be conducted on all spectrophotometers used in the ILS to demonstrate that they meet the proposed requirement.

9. Data Collection and Computation Requirements

9.1 The surrogate test method should specify exact conditions to be used in the collection of the spectral data. For example, the test method may specify some or all of the following:

9.1.1 Number of scans to be signal averaged or signal integration time,

9.1.2 Scan speed, and

9.1.3 Bandwidth, slit width or resolution.

9.2 If computations are required to convert the raw collected data to the form used by the multivariate model, the surrogate test method should specify exactly how those computations should be done. For example, for an FT-IR method, the surrogate test method should specify the type of apodization, level of zero-filling, and type of phase correction that is to be used in calculating the spectra.

9.3 If the surrogate test method requires that the spectral data be preprocessed prior to multivariate calibration or analysis, then the test method must specify exactly how the preprocessing is to be performed. The surrogate test method must mathematically define the preprocessing function including all parameters required for its computation either directly, or by reference to the literature. For example, if the second derivative of the spectrum must be calculated prior to multivariate calibration or analysis, then the test method must specify how the derivative is calculated. If, for instance, a Savitzky-Golay[3] digital filter is used, the test method should indicate which derivative, the polynomial degree and number of points for the digital filter and preferably list the digital filter parameters.

NOTE 6—Different types of preprocessing produce different multivariate models, and different analysis results. While the difference in some instances may be small, this cannot be assumed in developing a surrogate test method. If multiple types of preprocessing are to be allowed within a surrogate test method, then it is up to the test method developer to demonstrate that they all produce statistically indistinguishable results.

10. Recommended Limitations on Use of Multivariate Calculation Procedures

10.1 Typically, no two multivariate calibration models developed using differing algorithms or different numbers of variables will produce identical results. For example, PCR models built with differing numbers of principal components will typically show differences in their standard errors of calibration and qualification and relative biases in their predictions. Additionally, different models that appear to produce comparable results based on their standard errors of calibration and qualification may produce significantly different results when applied to actual samples. Therefore, it is strongly recommended that surrogate test methods employ only one discrete modeling procedure.

10.1.1 Do not assume that PLS and PCR produce equivalent models. Full-spectra surrogate test methods should specify either PLS or PCR, not both, unless PLS and PCR are shown to produce statistically indistinguishable results as discussed in 10.2.

10.1.2 The specific number of variables to be used in the model should be specified in the test method.

[3] Savitsky, A., and Golay, M.J., *Analytical Chemistry*, Vol 36, 1964, pp. 1627–1639, with corrections by Steiner, J., Termonia, Y., and Deltour, J., *Analytical Chemistry*, Vol 44, 1972, pp. 1906–1909.

10.1.3 Mean-centering, if used, should be a requirement, not an option. Autoscaling of spectral data should not be recommended for PCR or PLS calibrations. If autoscaling is employed for MLR calibrations, it should be a requirement, not an option.

NOTE 7—Autoscaling involves mean-centering the spectral data, and then scaling the data at each wavelength (frequency) by the standard deviation of the calibration set at that wavelength (frequency). For full-spectrum methods such as PCR and PLS, autoscaling can scale up the variance associated with noise occurring at spectral baseline points relative to the variance associated with signal at spectral features, thus effectively decreasing the signal-to-noise of the data.

10.2 If the test method developer wants to include more than one modeling algorithm or a range of numbers of variables, then the developer is responsible for demonstrating that all allowed modeling procedures produce statistically indistinguishable analyses when applied to the actual samples used in the ILS.

NOTE 8—The errors between different multivariate models developed using the same calibration data set are not completely independent. Thus, the statistical tests described in 10.2.1 – 10.3.2 thatassume independence, are not strictly applicable. The tests may allow a small number of models to pass as equivalent when there are in fact small biases or differences in precision, however, the tests are not expected to indicate inequality for models that are, in fact, equivalent.

10.2.1 Calibration and qualification data from each laboratory that participates in the ILS should be modeled using all proposed algorithms and ranges of variables. The various models should all be applied for analysis of the spectra of the ILS samples, and the concentration/property estimates from each model should be compared in terms of bias and precision.

10.2.2 For each proposed modeling procedure, calculate the surrogate test method repeatability using the data from the ILS. Compare all possible pairs of repeatability estimates using an *F*-test:

$$F_{repeatability} = \frac{r_i^2}{r_j^2} \text{ if } r_i > r_j \text{ , } F_{repeatability} = \frac{r_i^2}{r_i^2} \text{ if } r_i < r_j \quad (9)$$

where r_i and r_j are the calculated repeatability for two different multivariate modeling procedures. Compare $F_{repeatability}$ to the critical *F*-value from Table 2 where the degrees of freedom for both the numerator and denominator should both be the repeatability degrees of freedom from the ILS.

10.2.2.1 If $F_{repeatability}$ is less than the critical *F*-value, then the repeatability of the results produced by the two different multivariate modeling procedures are comparable. Continue with the reproducibility and bias tests.

10.2.2.2 If $F_{repeatability}$ is greater than the critical *F*-value, then the repeatability of the results produced by the two different multivariate modeling procedures are not comparable. The surrogate test method should not include both as multivariate modeling procedures.

10.2.3 For each proposed modeling procedure, calculate the surrogate test method reproducibility using the data from the ILS. Compare all possible pairs of repeatability estimates using an *F*-test:

$$F_{reproducibility} = \frac{R_i^2}{R_j^2} \text{ if } R_i > R_j \text{ , } F_{reproducibility} = \frac{R_j^2}{R_i^2} \text{ if } R_i < R_j \quad (10)$$

where R_i and R_j are the calculated reproducibility for two different multivariate modeling procedures. Compare $F_{reproducibility}$ to the critical *F*-value from Table 2.

10.2.3.1 If $F_{reproducibility}$ is less than the critical *F*-value, then the reproducibility of the results produced by the two different multivariate modeling procedures are comparable. Continue with the bias test.

10.2.3.2 If $F_{reproducibility}$ is greater than the critical *F*-value, then the reproducibility of the results produced by the two different multivariate modeling procedures are not comparable.

TABLE 2 97.5 Percentiles of the *F* Statistic (for Two-Sided Test)

Denominator, Degrees of Freedom	Numerator													
	7	8	9	10	12	14	16	18	20	25	30	40	50	100
7	4.99	4.90	4.82	4.76	4.67	4.60	4.54	4.50	4.47	4.40	4.36	4.31	4.28	4.21
8	4.53	4.43	4.36	4.30	4.20	4.13	4.08	4.03	4.00	3.94	3.89	3.84	3.81	3.74
9	4.20	4.10	4.03	3.96	3.87	3.80	3.74	3.70	3.67	3.60	3.56	3.51	3.47	3.40
10	3.95	3.85	3.78	3.72	3.62	3.55	3.50	3.45	3.42	3.35	3.31	3.26	3.22	3.15
11	3.76	3.66	3.59	3.53	3.43	3.36	3.30	3.26	3.23	3.16	3.12	3.06	3.03	2.96
12	3.61	3.51	3.44	3.37	3.28	3.21	3.15	3.11	3.07	3.01	2.96	2.91	2.87	2.80
13	3.48	3.39	3.31	3.25	3.15	3.08	3.03	2.98	2.95	2.88	2.84	2.78	2.74	2.67
14	3.38	3.29	3.21	3.15	3.05	2.98	2.92	2.88	2.84	2.78	2.73	2.67	2.64	2.56
15	3.29	3.20	3.12	3.06	2.96	2.89	2.84	2.79	2.76	2.69	2.64	2.59	2.55	2.47
16	3.22	3.12	3.05	2.99	2.89	2.82	2.76	2.72	2.68	2.61	2.57	2.51	2.47	2.40
17	3.16	3.06	2.98	2.92	2.82	2.75	2.70	2.65	2.62	2.55	2.50	2.44	2.41	2.33
18	3.10	3.01	2.93	2.87	2.77	2.70	2.64	2.60	2.56	2.49	2.44	2.38	2.35	2.27
19	3.05	2.96	2.88	2.82	2.72	2.65	2.59	2.55	2.51	2.44	2.39	2.33	2.30	2.22
20	3.01	2.91	2.84	2.77	2.68	2.60	2.55	2.50	2.46	2.40	2.35	2.29	2.25	2.17
25	2.85	2.75	2.68	2.61	2.51	2.44	2.38	2.34	2.30	2.23	2.18	2.12	2.08	2.00
30	2.75	2.65	2.57	2.51	2.41	2.34	2.28	2.23	2.20	2.12	2.07	2.01	1.97	1.88
35	2.68	2.58	2.50	2.44	2.34	2.27	2.21	2.16	2.12	2.05	2.00	1.93	1.89	1.80
40	2.62	2.53	2.45	2.39	2.29	2.21	2.15	2.11	2.07	1.99	1.94	1.88	1.83	1.74
45	2.58	2.49	2.41	2.35	2.25	2.17	2.11	2.07	2.03	1.95	1.90	1.83	1.79	1.69
50	2.55	2.46	2.38	2.32	2.22	2.14	2.08	2.03	1.99	1.92	1.87	1.80	1.75	1.66
60	2.51	2.41	2.33	2.27	2.17	2.09	2.03	1.98	1.94	1.87	1.82	1.74	1.70	1.60
70	2.47	2.38	2.30	2.24	2.14	2.06	2.00	1.95	1.91	1.83	1.78	1.71	1.66	1.56
80	2.45	2.35	2.28	2.21	2.11	2.03	1.97	1.92	1.88	1.81	1.75	1.68	1.63	1.53
90	2.43	2.34	2.26	2.19	2.09	2.02	1.95	1.91	1.86	1.79	1.73	1.66	1.61	1.50

The surrogate test method should not include both as multivariate modeling procedures.

NOTE 9—Although the *F*-tests in 10.2.2 and 10.2.3 are two-sided tests conducted at the 95 % probability level, the critical *F* value against which the calculated $F_{repeatability}$ and $F_{reproducibility}$ are compared come from the 97.5 percentiles of the *F*-statistic (Table 2). If the ratio r_a^2/r_b^2 (or R_a^2/R_b^2) was calculated without requiring that the larger variance be in the numerator, the calculated $F_{repeatability}$ ($F_{reproducibility}$) value would have to be compared against both the lower 2.5 percentile point and the upper 97.5 percentile point of the *F*-distribution to determine if the two variances were statistically distinguishable. Because of the nature of the *F*-distribution, comparing r_a^2/r_b^2 (or R_a^2/R_b^2) to the 2.5 percentile is equivalent to comparing r_b^2/r_a^2 (or R_b^2/R_a^2) to the 97.5 percentile point. Requiring that larger variance is always in the numerator allows the "two-tailed" test to be accomplished in one step. If the variance of the two populations were equal, then there would be only a 2.5 % chance that $r_a^2 > r_b^2$ by more than the tabulated amount, and a 2.5 % chance that $r_a^2 < r_b^2$ by more than the tabulated amount with degrees of freedom reversed.

10.2.4 For each sample used in the ILS, results for the proposed modeling procedures are compared pairwise. For each modeling procedure, calculate the grand average of the estimates over replicates from all *m* laboratories. Calculate the difference between these average values:

$$bias_i(a,b) = \frac{\sum_{j=1}^{l} \bar{y}_{ij}(a)}{m} - \frac{\sum_{j=1}^{l} \bar{y}_{ij}(b)}{m} \quad (11)$$

where $\bar{y}_{ij}$(a) and $\bar{y}_{ij}$(b) are means of the replicate estimates for the i^{th} sample measured in the j^{th} laboratory using modeling procedures *a* and *b* respectively. Calculate a *t*-value as

$$t = \frac{\sqrt{2d}\,|bias_i(a,b)}{\sqrt{(R_i(a)/2.77)^2 + (R_i(b)/2.77)^2}} \quad (12)$$

where:

$R_i(a)$ and $R_i(b)$ = the reproducibilities established from the ILS for results obtained using multivariate procedures *a* and *b* respectively, and

d = the reproducibility degrees of freedom used in calculating $R_i(a)$ and $R_i(b)$.

Compare the calculated *t*-value to the critical *t*-value in Table 3 for *d* degrees of freedom.

10.2.4.1 If the calculated *t*-value is less than the critical *t*-value for all samples in the ILS, then any bias between the results produced by the alternative multivariate modeling procedures is statistically insignificant.

TABLE 3 95th Percentile of Student's |t| Distribution

Degrees of Freedom	*t*
1	12.7062
2	4.3027
3	3.1824
4	2.7764
5	2.5706
6	2.4469
7	2.3646
8	2.3060
9	2.2622
10	2.2281
11	2.2010
12	2.1788
13	2.1604
14	2.1448
15	2.1314
16	2.1199
17	2.1098
18	2.1009
19	2.0930
20	2.0860
21	2.0796
22	2.0739
23	2.0687
24	2.0639
25	2.0595
26	2.0555
27	2.0518
28	2.0484
29	2.0452
30	2.0423
31	2.0395
32	2.0369
33	2.0345
34	2.0322
35	2.0301
36	2.0281
37	2.0262
38	2.0244
39	2.0227
40	2.0211
41	2.0195
42	2.0181
43	2.0167
44	2.0154
45	2.0141
46	2.0129
47	2.0117
48	2.0106
49	2.0096
50	2.0086
55	2.0040
60	2.0003
65	1.9971
70	1.9944
75	1.9921
80	1.99006
85	1.98827
90	1.98667
95	1.98525
100	1.98397
105	1.98282
110	1.98177
115	1.98081
120	1.97993
125	1.97912
130	1.97838
135	1.97769
140	1.97705
145	1.97646
150	1.97591
155	1.97539
160	1.97490
165	1.97445
170	1.97402
175	1.97361
180	1.97323
185	1.97287
190	1.97253
195	1.97220
200	1.97190

10.2.4.2 If the calculated *t*-value is greater than the critical *t*-value for any of the individual samples in the ILS, then the alternative multivariate modeling procedures produce results

that differ by a statistically significant amount. The surrogate test method should not include both as multivariate modeling procedures.

10.3 In many cases, it is desirable to compare the results produced by the surrogate test method to results produced by another analytical test method. A preferred multivariate modeling procedure may be chosen so as to produce the best agreement with the alternative test method. Procedures for comparing the surrogate and alternative test methods are beyond the scope of this practice.

10.3.1 If a choice among multivariate modeling procedures is to be made based on comparisons to an alternative analytical test method, then such comparisons should be done independently of the ILS used to establish the precision of the surrogate test method. The data from the ILS that establishes the precision should also be used to estimate the bias between the surrogate and alternative test methods.

10.3.2 If a comparison of the surrogate and alternative analytical test method is done, then the surrogate test method should report the results of that comparison in terms of a prediction equation that relates results from the surrogate test method (independent variable) to those of the alternative method (dependent variable), and the statistics associated with the prediction equation.

11. Keywords

11.1 fluorescence spectroscopy; infrared spectroscopy; molecular spectroscopy; multivariate analysis; quantitative analysis; Raman spectroscopy; spectrometer qualification; spectrophotometer qualification; ultraviolet-visible spectroscopy