Designation: E 2617 – 08a

Standard Practice for
Validation of Empirically Derived Multivariate Calibrations[1]

This standard is issued under the fixed designation E 2617; 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 covers requirements for the validation of empirically derived calibrations (Note 1) such as calibrations derived by Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Squares (PLS), Artificial Neural Networks (ANN), or any other empirical calibration technique whereby a relationship is postulated between a set of variables measured for a given sample under test and one or more physical, chemical, quality, or membership properties applicable to that sample.

NOTE 1—Empirically derived calibrations are sometimes referred to as "models" or "calibrations." In the following text, for conciseness, the term "calibration" may be used instead of the full name of the procedure.

1.2 This practice does not cover procedures for establishing said postulated relationship.

1.3 This practice serves as an overview of techniques used to verify the applicability of an empirically derived multivariate calibration to the measurement of a sample under test and to verify equivalence between the properties calculated from the empirically derived multivariate calibration and the results of an accepted reference method of measurement to within control limits established for the prespecified statistical confidence level.

1.4 *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:* [2]
E 131 Terminology Relating to Molecular Spectroscopy
E 1655 Practices for Infrared Multivariate Quantitative Analysis

E 1790 Practice for Near Infrared Qualitative Analysis

3. Terminology

3.1 For terminology related to molecular spectroscopic methods, refer to Terminology E 131. For terminology related to multivariate quantitative modeling refer to Practices E 1655. While Practices E 1655 is written in the context of multivariate spectroscopic methods, the terminology is also applicable to other multivariate technologies.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *accuracy*—the closeness of agreement between a test result and an accepted reference value.

3.2.2 *bias*—the arithmetic average difference between the reference values and the values produced by the analytical method under test, for a set of samples.

3.2.3 *detection limit*—the lowest level of a property in a sample that can be detected, but not necessarily quantified, by the measurement system.

3.2.4 *estimate*—the constituent concentration, identification, or other property of a sample as determined by the analytical method being validated.

3.2.5 *initial validation*—validation that is performed when an analyzer system is initially installed or after major maintenance.

3.2.6 *Negative Fraction Identified*—the fraction of samples not having a particular characteristic that is identified as not having that characteristic.

3.2.6.1 *Discussion*—Negative Fraction Identified assumes that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.7 *ongoing periodic revalidation*—the quality assurance process by which, in the case of quantitative calibrations, the bias and precision or, in the case of qualitative calibrations, the Positive Fraction Identified and Negative Fraction Identified performance determined during initial validation are shown to be sustained.

3.2.8 *Positive Fraction Identified*—the fraction of samples having a particular characteristic that is identified as having that characteristic.

3.2.8.1 *Discussion*—Positive Fraction Identified assumes that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.9 *precision*—the closeness of agreement between independent test results obtained under stipulated conditions.

3.2.9.1 *Discussion*—Precision may be a measure of either the degree of reproducibility or degree of repeatability of the analytical method under normal operating conditions. In this context, reproducibility refers to the use of the analytical procedure in different laboratories, as in a collaborative study.

3.2.10 *quantification limit*—the lowest level of a sample property which can be determined with acceptable precision and accuracy under the stated experimental conditions.

3.2.11 *range*—the interval between the upper and lower levels of a property (including these levels) that has been demonstrated to be determined with a suitable level of precision and accuracy using the method as specified.

3.2.12 *reference value*—the metric of a property as determined by well-characterized method, the accuracy of which has been stated or defined, that is, another, already-validated method.

3.2.13 *validation*—the statistically quantified judgment that an empirically derived multivariate calibration is applicable to the measurement on which the calibration is to be applied and can perform property estimates with, in the case of quantitative calibrations, acceptable precision, accuracy and bias or, in the case of qualitative calibrations, acceptable Positive Fraction Identified and Negative Fraction Identified, as compared with results from an accepted reference method.

4. Summary of Practice

4.1 Validating an empirically derived multivariate calibration (model) consists of four major procedures: validation at initial development, revalidation at initial deployment or after a revision, ongoing periodic revalidation, and qualification of each measurement before using the calibration to estimate the property(s) of the sample being measured.

5. Significance and Use

5.1 This practice outlines a universally applicable procedure to validate the performance of a quantitative or qualitative, empirically derived, multivariate calibration relative to an accepted reference method.

5.2 This practice provides procedures for evaluating the capability of a calibration to provide reliable estimations relative to an accepted reference method.

5.3 This practice provides purchasers of a measurement system that incorporates an empirically derived multivariate calibration with options for specifying validation requirements to ensure that the system is capable of providing estimations with an appropriate degree of agreement with an accepted reference method.

5.4 This practice provides the user of a measurement system that incorporates an empirically derived multivariate calibration with procedures capable of providing information that may be useful for ongoing quality assurance of the performance of the measurement system.

5.5 Validation information obtained in the application of this practice is applicable only to the material type and property range of the materials used to perform the validation and only for the individual measurement system on which the practice is completely applied. It is the user's responsibility to select the property levels and the compositional characteristics of the validation samples such that they are suitable to the application. This practice allows the user to write a comprehensive validation statement for the analyzer system including specific limits for the validated range of application and specific restrictions to the permitted uses of the measurement system. Users are cautioned against extrapolation of validation results beyond the material type(s) and property range(s) used to obtain these results.

5.6 Users are cautioned that a validated empirically derived multivariate calibration is applicable only to samples that fall within the subset population represented in the validation set. The estimation from an empirically derived multivariate calibration can only be validated when the applicability of the calibration is explicitly established for the particular measurement for which the estimation is produced. Applicability cannot be assumed.

6. Methods and Considerations

6.1 When validating an empirically derived multivariate calibration, it is the responsibility of the user to describe the measurement system and the required level of agreement between the estimations produced by the calibration and the accepted reference method(s).

6.2 When validating a measurement system incorporating an empirically derived multivariate calibration, it is the responsibility of the user to satisfy the requirements of any applicable tests specific to the measurement system including any Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ) requirements; which may be mandated by competent regulatory authorities, an applicable Quality Assurance (QA), or Standard Operating Procedure (SOP) or be recommended by the instrument or equipment manufacturer.

6.3 *Reference Values and Quality Controls for the Accepted Reference Method*:

6.3.1 The reference (or true) value which is compared with each respective estimate produced by the empirically derived multivariate calibration is established by applying an accepted reference method, the characteristics of which are known and stated, to the sample from which the measurement system derives the measurement.

6.3.2 To ensure the reliability of the reference values provided by an accepted reference method, appropriate quality controls should be applied to the accepted reference method.

7. Procedure

7.1 The objective of the validation procedure is to quantify the performance of an empirically derived multivariate calibration in terms of, in the case of quantitative calibrations, precision, accuracy and bias or, in the case of qualitative calibrations, Positive Fraction Identified and Negative Fraction Identified relative to an accepted reference method for each property of interest. The user must specify, based on the intended use of the calibration, acceptable precision and bias or Positive Fraction Identified and Negative Fraction Identified performance criteria before initiating the validation. These criteria will be dependent on the intended use of the analyzer and may be based, all or in part, on risk based criteria.

7.1.1 The acceptable performance criteria specified by the user may be constant over the entire range of sample variability. Alternatively, different acceptable performance criteria may be specified by the user for different sub-ranges of the full sample variability.

7.2 Validation of calibration is accomplished by using the calibration to estimate the property(s) of a set of validation samples and statistically comparing the estimates for these samples to known reference values. Validation requires thorough testing of the model with a sufficient number of representative validation samples to ensure that it performs adequately over the entire range of possible sample variability.

7.3 *Initial Validation Sample Set* :

7.3.1 For the initial validation of a multivariate model, an ideal validation sample set will:

7.3.1.1 Contain samples that provide sufficient examples of all combinations of variation in the sample properties which are expected to be present in the samples which are to be analyzed using the calibration;

7.3.1.2 Contain samples for which the ranges of variation in the sample properties is comparable to the ranges of variation expected for samples that are to be analyzed using the model;

7.3.1.3 Contain samples for which the respective variations of the sample properties are uniformly and mutually independently distributed over their full respective ranges or, when applicable, subranges of variation; and

7.3.1.4 Contain a sufficient number of samples to statistically test the relationships between the measured variables and the properties that are modeled by the calibration.

7.3.2 For simple systems, sufficient validation samples can generally be obtained to meet the criteria in 7.3.1.1-7.3.1.4. For complex mixtures, obtaining an ideal validation set may be difficult if not impossible. In such cases, it may be necessary to validate discrete subranges of the calibration incrementally, over time as samples become available.

7.3.3 The number of samples needed to validate a calibration depends on the complexity of the calibration, the ranges of property variation over which the calibration is to be applied, and the degree of confidence required. It is important to validate a calibration with as many samples as possible to maximize the likelihood of challenging the calibration with rarely occurring, but potentially troublesome samples. The number and range of validation samples should be sufficient to validate the calibration to the statistical degree of confidence required for the application. In all cases, a minimum of 20 validation samples is recommended. In addition, the validation samples should:

7.3.3.1 Multivariately span the ranges of sample property values over which the calibration will be used; that is, the span and the standard deviation of the ranges of sample property values for the validation samples should be at least 100 % of the spans of the sample property values over which the calibration will be used, and the sample property values for the validation samples should be distributed as uniformly as possible throughout their respective ranges, and the variations of the sample property values among the samples should be as mutually independent as possible; and

7.3.3.2 Span the ranges of the independent variables over which the calibration will be used; that is, if the range of an independent variable is expected to vary from *a* to *b*, and the standard deviation of the independent variable is *c*, then the variations of that independent variable in the set of validation samples should cover at least 100 % of the range from *a* to *b*, and should be distributed as uniformly as possible across the range such that the standard deviation in that independent variable estimated for the validation samples will be at least 95 % of *c*.

(1) When validating a calibration for which detection limit or quantification limit is an important consideration, the user should include a number of validation samples whose property(s) are close to the detection or quantification limit(s) sufficient to validate the respective limit(s) to the statistical degree of confidence required for the application.

7.4 For quantitative calibrations, the validation error for each property in each sample is given by the Standard Error of Validation (SEV) and bias for that property.

7.4.1 The validation bias, $\bar{e}_v$, is a measure of the average difference between the estimates made based on the empirical model and the results obtained on the same validation samples using the reference method.

7.4.1.1 If there are single reference values and estimates for each validation sample, the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v}(\hat{v}_i - v_i)}{v} \tag{1}$$

where:
$\hat{v}_i$ = estimate from the model for the *i*th sample,
v_i = accepted reference value for the *i*th sample, and
v = number of validation samples.

7.4.1.2 If replicate estimates and a single reference value are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_{ij} - v_i)}{\sum_{i=1}^{v}r_i} \tag{2}$$

where:
$\hat{v}_{ij}$ = the *j*th estimate for the *i*th validation sample, and
r_i = number of replicate estimates for the *i*th validation sample.

7.4.1.3 If a single estimate and multiple reference values are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(\hat{v}_i - v_{ij})}{\sum_{i=1}^{v}s_i} \tag{3}$$

where:
$\hat{v}_i$ = estimate for the *i*th validation sample,
v_{ij} = the *j*th reference value for the *i*th validation sample, and

s_i = number of replicate reference values for the ith validation sample.

7.4.1.4 If multiple estimates and multiple reference values are available for the validation samples, then the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}\sum\limits_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik})}{\sum\limits_{i=1}^{v} r_i s_i} \tag{4}$$

where:
$\hat{v}_{ij}$ = the jth estimate for the ith validation sample,
v_{ik} = the kth reference value for the ith validation sample,
r_i = number of replicate estimates for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

7.4.2 The SEV, also called the Standard Error of Prediction (SEP) and the Standard Deviation of Validation Residuals (SDV), are measures of the expected agreement of the empirical model and the reference method. The calculation of SEV and SDV depend on whether replicate estimates or reference values, or both, are used.

7.4.2.1 If there are single reference values and estimates for each validation sample, then SEV and SDV are calculated as:

$$SEV = \sqrt{\frac{\sum\limits_{i=1}^{v}(\hat{v}_i - v_i)^2}{v}} \tag{5}$$

$$SDV = \sqrt{\frac{\sum\limits_{i=1}^{v}(\hat{v}_i - v_i - \bar{e}_v)^2}{v}}$$

where:
$\hat{v}_i$ = estimate from the model for the ith sample,
v_i = accepted reference value for the ith sample, and
v = number of validation samples.

7.4.2.2 If replicate estimates and a single reference value are available for the validation samples, then SEV and SDV are calculated as:

$$SEV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}(\hat{v}_{ij} - v_j)^2}{\sum\limits_{i=1}^{v} r_i}} \tag{6}$$

$$SDV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}(\hat{v}_{ij} - v_i - \bar{e}_v)^2}{\sum\limits_{i=1}^{v} r_i}}$$

where:
$\hat{v}_{ij}$ = the jth estimate for the ith validation sample, and
r_i = number of replicate estimates for the ith validation sample.

NOTE 2—If each validation sample is estimated r times, an average estimate could be used in 7.4.2.1, but then the SEV calculated would represent the expected agreement between the average of r estimations and a single reference measurement, not the agreement based on a single estimation from the empirical model.

7.4.2.3 If a single estimate and multiple reference values are available for the validation samples, then SEV and SDV are calculated as:

$$SEV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}(\hat{v}_i - v_{ij})^2}{\sum\limits_{i=1}^{v} s_i}} \tag{7}$$

$$SDV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}(\hat{v}_i - v_{ij} - \bar{e}_v)^2}{\sum\limits_{i=1}^{v} s_i}}$$

where:
$\hat{v}_i$ = estimate for the ith validation sample,
$\hat{v}_{ij}$ = the jth reference value for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

NOTE 3—If each validation sample has s reference values, an average estimate could be used in 7.4.2.1, but then the SEV calculated would represent the expected agreement between an estimate from the empirical model and the average of s reference measurements, not a the agreement relative to a single reference measurement.

7.4.2.4 If multiple estimates and multiple reference values are available for the validation samples, then SEV and SDV are calculated as:

$$SEV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}\sum\limits_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik})^2}{\sum\limits_{i=1}^{v} r_i s_i}} \tag{8}$$

$$SDV = \sqrt{\frac{\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{r_i}\sum\limits_{k=1}^{s_i}(\hat{v}_{ij} - v_{ik} - \bar{e}_v)^2}{\sum\limits_{i=1}^{v} r_i s_i}}$$

where:
$\hat{v}_{ij}$ = the jth estimate for the ith validation sample,
$\hat{v}_{ik}$ = the kth reference value for the ith validation sample,
r_i = number of replicate estimates for the ith validation sample, and
s_i = number of replicate reference values for the ith validation sample.

NOTE 4—If each validation sample has r estimates and s reference values, average estimates and reference values could be used in 7.4.1.1, but then the SEV calculated would represent the expected agreement between r estimates from the empirical model and the average of s reference measurements, not a the agreement between a single estimate and reference measurement.

7.4.3 *Significance of Validation Bias*— A t-value can be calculated as:

$$t = \frac{e_v d_v}{SDV} \tag{9}$$

where:
d_v = degrees of freedom and is equal to the denominator in the bias calculation.

NOTE 5—The t-value is compared to a critical t-value for the desired

probability level (typically 95 %).

7.4.3.1 If the calculated *t*-value is less than the critical *t*-value, then the validation bias is not statistically significant and the empirical model and reference method are expected to on average yield the same result. In this case, either SEV or SDV are adequate measures of the expected agreement between the empirical model and the reference method. If the validation bias is of practical significance relative to the user specified bias requirement, then the precision of the empirical model results is insufficient to achieve the user requirement.

7.4.3.2 If the calculated *t*-value is greater than the critical *t*-value, then the validation bias is statistically significant. In this case SDV is a better measure of the expected agreement between the results of the empirical model and the reference method. While the bias may be statistically significant, it may not be of practical significance relative to the user specified requirements for the empirical model.

7.5 *Positive and Negative Fractions Identified*:

7.5.1 The Positive Fraction Identified of the calibration is given by: Positive Fraction Identified = (number of samples identified as having a stated characteristic) / (total number of samples having the stated characteristic).

7.5.2 The Negative Fraction Identified of the calibration is given by: Negative Fraction Identified = (number of samples identified as not having a stated characteristic) / (total number of samples not having the stated characteristic).

7.5.3 The equations for Positive Fraction Identified and Negative Fraction Identified assume that the characteristic being measured either is or isn't present. It is not applicable to tests with multiple possible outcomes.

7.6 The users should use statistical tests and decision criteria appropriate to the application to decide if the SEV and bias are within statistically acceptable limits.

7.7 *Samples for Revalidation After Initial Deployment and Ongoing Periodic Revalidation Samples*:

7.7.1 The user must determine, based on the particulars of each application, the appropriate timing and number of samples required for revalidation after initial deployment and for ongoing periodic revalidation.

7.7.1.1 The timing and number of revalidation samples may be adjusted from time to time as experience is gained in applying the calibration under actual conditions.

7.7.1.2 In many cases revalidation samples are restricted to "samples of opportunity" and limited to samples from actual production operations. In such cases, care should be taken to schedule revalidation samples as asynchronously as possible with respect to recurring conditions such as time of day, production process operating conditions, phase or stage of production process, ambient conditions, operating personnel, etc. This listing of potential conditions for consideration is exemplary, not comprehensive; the user should take into account any external conditions pertinent to the application.

7.7.2 It is recommended that the results of ongoing periodic revalidation should be monitored or tracked by control charting.

8. Qualification of Each Measurement Prior to Application of the Validated Calibration

8.1 The independent variables measured from a sample under test must be evaluated to ensure that this measurement is qualified to be processed by the calibration to produce estimates of the property(s) of interest. The purpose of this qualification step is to determine, within user specified statistical limits, if the validation samples used to validate the calibration are sufficiently representative of (similar to) the sample under test. If the measurement is qualified, the estimates should fall within accuracy and precision bounds determined during the validation. If the measurement is not qualified, then the accuracy and precision of the estimates are not known based on the validation. The measurement of a sample under test may be qualified using Mahalanobis distance, Nearest Neighbor Mahalanobis Distance (NNMD), or Standard Residual Variance in the Independent Variables (SRVIV), either singly or in combination. The user may also specify additional qualification criteria if and as appropriate to the application.

8.1.1 The development of an empirical model will typically involve transformation of the independent variables. By way of illustration, such transformation may include one or more of the following:

8.1.1.1 Linearization of the independent variables (for example, conversion from transmission to absorbance, from reflectance to log(1-reflectance), etc.);

8.1.1.2 Digital filtering (smoothing, digital derivatives);

8.1.1.3 Orthogonalization (Orthogonal Signal Correction);

8.1.1.4 Rank reduction (Principal Components Analysis (PCA) or PLS);

8.1.1.5 Squares, cross products or nonlinear functions of variables;

8.1.1.6 Explicit artifact removal (cosmic ray event removal);

8.1.1.7 Centering or baseline correction;

8.1.1.8 Arbitrary scaling, variance scaling, or auto scaling;

8.1.1.9 Exclusion of one or more independent variables from use in the calibration; and

8.1.1.10 Integration of peaks with or without baseline correction.

8.1.2 Mahalanobis distance, NNMD, and SRVIV statistics are calculated after applying the same transformations to the measurement being qualified which were applied to the measurements used to produce and validate the calibration.

8.2 SRVIVs can sometimes be employed to determine if the samples used to validate the empirical model are sufficiently representative of (similar to) the sample under test. SRVIV is intended to detect any anomalous variance which may be present in the measurement from new signals (for example, new chemical components, new instrumental or sample conditions, etc.) that were not represented in the validation samples. If the validation samples are sufficiently representative of the (unknown) sample under test, then the amount of residual variance in the independent variables of the sample under test will be statistically indistinguishable from the amount of residual variance in the validation samples. This is always a necessary criterion for qualification testing, but it may not

always be solely sufficient. If the empirical calibration utilizes most of the non-noise portion of variance in the independent variables, the residual variance will be a very sensitive measure of any aberrant variance present in the data for the sample under test. Alternatively, if the empirical model is based on a small fraction of the non-noise portion of the variance in the independent variable, then tests based on the statistics of the SRVIV are unlikely, used alone, to provide adequate warning of measurements, which are not qualified for estimation by the calibration.

8.2.1 The residual variance in the independent variables is defined as that fraction of the variance in the variables which is not spanned by the validation samples' basis space comprising an appropriate number of abstract factors determined by either PCA (1, 2)[3] or PLS (1, 2). If the matrix P comprises column vectors, each of which represents one of the factors comprising the PCA basis space, the matrix A comprises column vectors, each of which contains the independent variables (for example, the spectrum) of a validation sample, and the matrix R comprises column vectors, each of which is the residual variance in each corresponding spectrum in A which is not spanned by the basis space, then:

$$R = A - (A^T P P^T)^T \qquad (10)$$

and the standard residual (S.R.) is then given by:

$$S.R. = \sqrt{\frac{\sum_{m=0}^{r} \sum_{n=0}^{c} R_{m,n}^2}{r \times c - p - 1}} \qquad (11)$$

where:
r = number of rows in A,
c = number of columns in A,
p = number of columns in P, and
$(r \times c - p - 1)$ = number of degrees of freedom of the residuals.

8.2.1.1 For the purposes of this calculation, the user should determine that all rows of the matrix of residuals have equal variance and that all columns of the matrix of residuals have equal variance before applying the equation for standard residual.

8.2.1.2 If the abstract factors for the basis space were calculated using PLS, Eq 11 is an approximation in that it does not account for the fact that, in general, each additional PLS factor does not reduce the degrees of freedom of the system by exact integer amount.

8.2.1.3 PLS algorithms, depending upon how they are implemented in software, can produce either orthogonal PLS factors (often called the PLS loading weights) or non-orthogonal PLS factors (often called the PLS loadings). Either type of factors may be used for the basis space, but each type of factor may yield different qualification results. The orthogonal PLS loading weights will usually produce a standard residual for the validation samples which is very close in magnitude to the standard residual produced by the corresponding number of PCA factors. The non orthogonal PLS

loadings will usually produce a standard residual for the validation samples which is slightly or significantly larger than the standard residual produced by the corresponding number of PCA factors. Accordingly, the SRVIV test when using the PLS loading weights may provide more sensitive detection of unqualified samples in some cases.

8.2.2 After the S.R. is calculated for the validation samples, a user specified confidence limit is then applied to the S.R. to establish a range of expected S.R. values for samples that are well represented by the validation samples. Based on the user specified confidence limit users establish an upper cutoff for the S.R. for samples under test.

8.2.3 If the column vector, B, contains the measurement of the sample under test, the residual, R, of its independent variables is given by:

and the standard residual, S.R., is then given by:

$$R = B - (B^T P P^T)^T \qquad (12)$$

and the standard residual ($S.R.$) is then given by:

$$S.R. = \sqrt{\frac{\sum R^2}{r - 1}} \qquad (13)$$

where:
r = number of rows in B, and
$r - 1$ = number of degrees of freedom of the residuals.

8.2.4 If the S.R. value for the sample under test is less than the cutoff value established by the user in accordance with 8.2.2, then the measurement of the sample under test is not disqualified, with respect to residual variance in the independent variables, for estimation by the calibration. In such cases, the validation samples are sufficiently representative of the sample under test that estimates should fall within the accuracy and precision bounds determined during the validation.

8.2.5 If the S.R. value for the sample under test is greater than the user established cutoff value, then the measurement of the sample under test is not qualified, with respect to residual variance in the independent variables, for estimation by the calibration. Validation samples are not sufficiently representative of the sample under test and the accuracy and precision of the estimates are not known based on the validation.

8.3 The Mahalanobis distance can indicate whether or not the transformed variables for a sample under test fall within the multivariate space defined by the transformed variables for the validation sample set such that the sample under tests is an interpolated rather than an extrapolated sample with respect to the samples in the validation set. If the validation samples are sufficiently representative of the sample under test, then the Mahalanobis distance of the measurement with respect to the centroid of the validation samples will be statistically indistinguishable from the Mahalanobis distances of the set of validation samples.

8.3.1 The Mahalanobis distance, h, of the sample under test is given by:

$$h = x^T (X X^T)^{-1} x \qquad (14)$$

where:
x = the mean centered vector containing the measured values for the sample under test, and

---

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.