APPLYING MEASUREMENT SCIENCE TO ENSURE END ITEM PERFORMANCE[1]

Howard Castrup, Ph.D.
Integrated Sciences Group
14608 Casitas Canyon Rd
Bakersfield, CA 93306
Phone: 661-872-1683
Fax: 661-872-3669
hcastrup@isgmax.com

Abstract

A methodology is presented for the cost-effective determination of appropriate measurement assurance standards and practices for the management of calibration and test equipment. The methodology takes an integrated approach in which quality, measurement reliability and costs at each level in a test and calibration support hierarchy are linked to their counterparts at every other level in the hierarchy. This relates requirements and capabilities at any given level in the hierarchy to the performance objectives of the end items which the hierarchy is ultimately employed to support. Included in the methodology is the capability to quantify the effect of calibration/test support on costs resulting from the risk of degraded end item performance, either in terms of losses suffered through poor performance or expenses incurred from returned products, warranty rework or reimbursement, legal damages, or retrofit of product improvements.

INTRODUCTION

Traditionally, test and calibration, from the calibration of primary standards to the testing of end items have been focused primarily on workload, with no clear-cut relationship between the relevance of test or calibration results to the requirements of equipment users. Accordingly, each level of support has essentially been an isolated function, connected only by a thin thread of traceability and arbitrary nominal accuracy ratios or measurement decision risk criteria. At present, there are little means of establishing the optimal support required from each level of support to ensure acceptable performance of end items. Likewise, there is no widely recognized rigorous methodology for relating the development of end item specifications to available test and calibration support.

This paper describes an approach, labeled the "ETS model" [1, 2] for applying measurement science and cost modeling to establish functional links from level to level with the aim of ensuring acceptable end item performance in a cost-effective manner. The development of such links involves the application of uncertainty analysis, risk analysis, calibration interval analysis, life cycle cost analysis and end item utility modeling. It also involves the acquisition, interpretation, use and development of equipment attribute specifications.

With the exception of life cycle cost analysis, NCSLI recommended practices (RPs) are either available or are under development in each of these areas. These RPs include significant advances in methodology since the inception of the ETS model. The application of these advances to refining and expanding the model and suggestions for improving its life cycle cost analysis capability are discussed. Data requirements are also discussed.

The Test and Calibration Support Hierarchy

The overall infrastructure intended to provide measurement assurance for the testing of end items is called the "test and calibration support hierarchy." A simplified diagram of this hierarchy is shown in Figure 1. In the hierarchy, the end item is placed at the end of the chain. Prior to the end item is the test system and prior to the test system is a series of calibration functions, lumped under the heading "Calibration Standard" and culminating with the calibration function performing the calibration of the test system.

---

[1] Presented at the Measurement Science Conference, Anaheim, 2008.

Test and calibration hierarchies are characterized by technical and management parameters such as calibration system, test system and end item performance tolerances; calibration system and test system calibration intervals; accuracy ratios between calibration systems and test systems and between test systems and end items; measurement reliability (percent in-tolerance) targets; acceptable levels of measurement decision risk; equipment renewal policies; and so on.

Over the past three decades, the problem of analyzing and evaluating test and calibration technical and management parameters has received considerable study [3]. The results of this work have been expressed in various papers and technical reports in the form of tabulated or plotted comparisons and trends. While useful in the specific applications treated by each author, these tables and plots are usually of little value in general application. This is because the specifics of a given test and calibration support scenario typically take on a multitude of possible combinations obviate the use of a limited set of tables or plots.

The development of the theory and practice of test and calibration has been traditionally based on the now classical concept of straightforward comparisons between simple devices, each characterized by a single measurement attribute. Prior to the late `60s, the accuracy of higher-level devices or standards was generally significantly greater than the devices being tested or calibrated. Under such idyllic circumstances, all that was required for ensuring adequacy of testing or calibration was to legislate that only standards could be employed whose accuracies exceeded some easily attainable minimum level relative to the device under test.

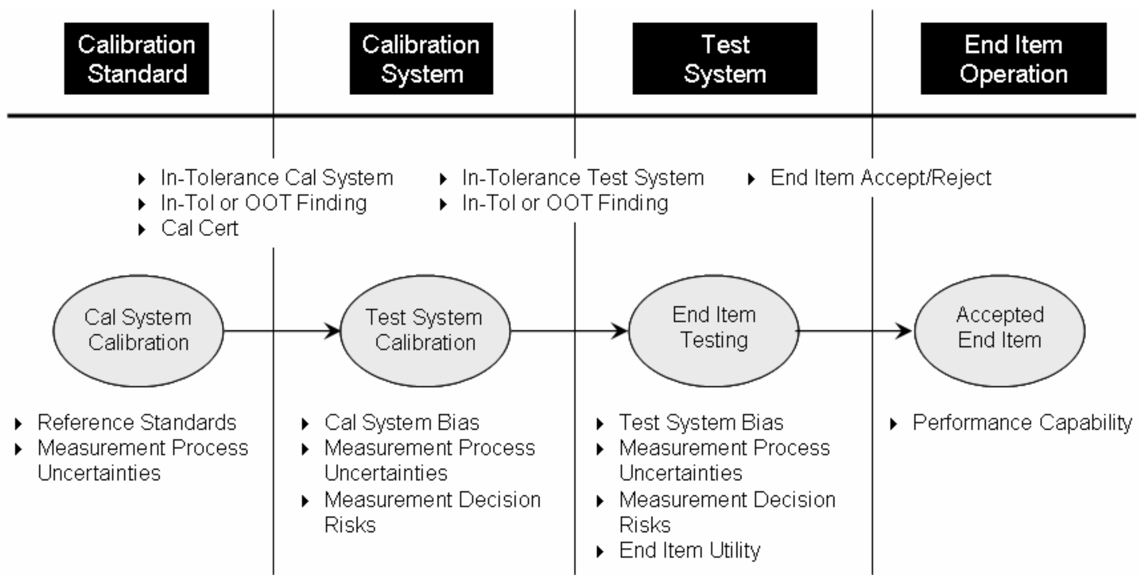

Figure 1. The test and calibration hierarchy. An end-to-end view of calibration and test support to end item performance. "End Item Operation" refers to the performance of end items accepted by testing. The testing function is represented by the "Test System" level. The utility of accepted end items is a function of the bias in the test system and the measurement uncertainties of the test process. This utility is also impacted by measurement decision risk, discussed later in this paper. The bias of the test system is a random variable whose distribution is affected by the accuracy of the "Calibration System." This is a function of the bias of the calibrating reference, the measurement uncertainties of the calibration process and by measurement decision risk. Finally, the bias of the calibration system is a random variable whose distribution is affected by the accuracy of its calibration by the "Calibration Standard." This is a function of the bias of the reference standard(s) and the measurement uncertainties of the calibration process.

In the modern technological environment, the classical concept of test and calibration is no longer generally applicable. The pressures of the competitive international marketplace and of national aerospace, energy, environmental and defense performance capability and reliability requirements have led to a situation in which end item performance tolerances rival and, in some cases, exceed the best accuracies attainable at the primary standards level. In such cases, the interpretation of test and calibration data and the management of test and calibration systems requires that the subtleties of the test/calibration process be accounted for and their impact on end item performance and support costs be quantified.

This argues for the development of tools by which a non-specialist can enter information defining his or her specific infrastructure, provide relevant technical and cost data, punch a button and receive technical and management parameter values and other decision support outputs that can be used to evaluate and optimize the infrastructure.

A PC DOS application was developed in the late '80s to determine whether such a tool was feasible [1]. The application incorporates a number of concepts and methods which emerged from analytical metrology research up to that time. Because it can be used as an analytical tool for defining and adjusting calibration, test and end item tolerances, among other things, it was named the Equipment Tolerancing System (ETS).

The ETS model employs analytical metrology to integrate the levels of the hierarchy by describing each level in terms of the support it receives from the next highest level and the support it gives to the next lowest level.

Testing of a given end item attribute by the test system yields a reported in-or out-of-tolerance indication (referenced to the attribute's tolerance limits), an adjustment (referenced to the attribute's adjustment limits) and a "post-test" measurement reliability (referenced to the attribute's tolerance limits).   Similarly, the results of calibration of the test system attribute are a reported in- or out-of-tolerance indication (referenced to the test system attribute's test limits), an attribute adjustment (referenced to the attribute's adjustment limits) and a beginning-of-period (BOP) measurement reliability (referenced to the attribute's performance limits). The same sort of data result from the calibration of the calibration system and accompany calibrations up through the hierarchy until a point is reached where the item calibrated is itself a calibration standard.

Ordinarily, calibration standards are not managed to specified performance or test tolerances and reported as in- or out-of-tolerance per se, but instead receive a reported measured value, accompanied by confidence limits. Since such standards are not managed to specified tolerances, a statement of BOP measurement reliability is apparently not applicable. In addition, the treatment of calibration standards differs from that of calibration or test systems since the calibration standards' measurement attribute values are reported rather than adjusted.

These observations appear to set the calibration of standards apart from other test or calibration scenarios. With regard to reported attribute values in place of adjustments, however, such reports can be considered to be completely equivalent to non-intrusive adjustments to center spec since reported values are used as nominal values until the next calibration. Additionally, the lack of specified tolerances for calibration standards will likely be eliminated in future calibration standard management systems. This is due to the fact that such standards are assigned calibration intervals, which can be optimized only if specified tolerances accompany reports of calibration. Specifically, the calibration standard attribute's reported measured value needs to be accompanied by both a set of limits (i.e., performance specifications), which are expected to contain the attribute value over the course of the calibration interval, and an estimate of the probability that this expectation will be realized (i.e., a measurement reliability target).

With regard to the applicability of the reliability analysis methodology employed in the ETS model, test or calibration at any pair of successive levels in the hierarchy is equivalent to test or calibration at any other pair of successive levels. This is not true, however, for cost modeling and analysis. For cost modeling, end item testing is treated somewhat differently than other tests or calibrations in that the cost consequences of accepting end items as having passed testing can be evaluated in intrinsic terms. This will be elaborated on later.

End-to-End Analysis

End-to-End Testing

End-to-end testing is emerging as a key element in the verification of operational systems [4]. The objective of end-to-end testing is to demonstrate interface compatibility and desired total functionality among different elements of a system and between systems, if applicable.

End-to-end testing includes executing complete operational scenarios for all possible configurations to ensure that all performance requirements are verified. It is employed during hardware and software selection and is applied throughout all phases of the system life cycle.

Applying Measurement Science to End-to-End Analysis

End-to-end analysis can be applied to the test and calibration hierarchy by regarding the hierarchy as a "system," with each level comprising a system module. Each module is characterized by a measurement area (e.g., AC voltage, length, mass, etc.), selected instrumentation, a statistical measurement error distribution and a measurement process uncertainty. The output of each module comprises an input to the next and influences its subsequent output. The final output of the system is a statistical distribution for values of a specific end item attribute that governs the performance capability of an end item function or feature.

The end-to-end concept was first applied to the test and calibration hierarchy in the ETS model. Technical elements in the end-to-end system analysis of the ETS model are

- The end item attribute performance requirements
- The end item attribute utility function
- The end item attribute bias distribution
- Test system error distributions
- Calibration system error distributions
- Calibration standards error distributions
- Test and calibration system calibration intervals
- Test and calibration system measurement reliability functions.

In addition to technical elements, the ETS model includes considerations of cost variables such as operational costs and the cost of end item performance capability, or lack thereof. The former include equipment life cycle costs and the latter are expressed is terms of the utility of the supported end item attribute. This and related concepts will be discussed later under cost optimization.

The ETS cost elements are

- Cost of test, calibration and repair
- Cost of rejection (with consequent adjustment, repair or rework and downtime) of in-tolerance UUT
- Cost of acceptance of the supported end item
- Equipment availability

Several important benefits accrue as a result of applying the ETS model. These include

- Elimination of arbitrary nominal TAR/TUR criteria (e.g., 4:1, the "ten percent rule," etc.)
- Elimination of arbitrary measurement decision risk requirements (e.g., 2%)
- The ability to make test and calibration support decisions on a return-on-investment basis
    - Equipment purchases
    - Software purchases
    - Calibration interval adjustments and reliability targets
    - Provision for spares
    - Personnel
    - Facilities

END ITEM PERFORMANCE

The Performance Requirement

The performance of an end item function can usually be linked to the value of one or more attributes. For example, the range of a cannon can be linked to the bore of the cannon and the diameters of cannonballs. In establishing this link, several variables come into play, as shown in Table 1. These variables are required to express the performance of the end item function in terms of the values of its associated attributes. The relationship between performance and attribute value is called the *utility function*.

Table 1. End Item Performance Variables

Variable Description	Variable Name
End item attribute value	x
End item attribute value corresponding to the onset of degraded performance	x_d
End item attribute value corresponding to loss of function	x_f
Successful end item attribute performance given x	$P(\text{success}\|x)$
Probability of a successful outcome, given successful end item performance	P_{sr}
Time elapsed since end item testing	t
Probability density function for the end item attribute of interest	$f(x,t)$
Expected end item life cycle	T

The Utility Function

In the ETS model, the values x_d and x_f mark, respectively, the onset of degraded performance and the complete loss of utility. The ETS utility function is given by

$$P(\text{success} \mid x) = \begin{cases} 1, & |x| \le x_d \\ \cos^2\left[\dfrac{(|x| - x_d)\pi}{2(x_f - x_d)}\right], & x_d \le |x| \le x_f \\ 0, & x_f \le |x|, \end{cases} \tag{1}$$

where $P(\text{success}|x)$ is the probability for successful performance, given that the attribute value is equal to x. The form of this utility function is shown in Figure 2. The relationship between attribute value and performance is shown in Figure 3.

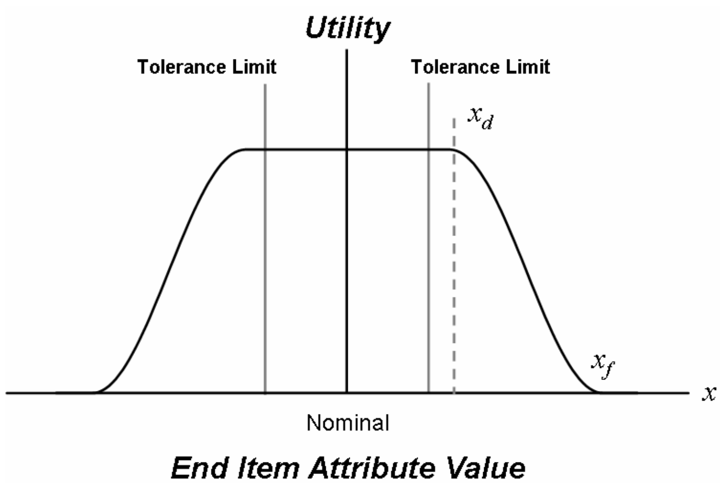

Figure 2. **The End Item Utility Function**. In the ETS model, the utility of an end item performance characteristic is a function of the value x of a specific attribute. In the case shown,[2] the attribute tolerance limits are located inside the range $-x_d$ to x_d. Cases where the limits lie outside this range are marked by less rigorous attribute specifications. More will be said on this later.

[2] Taken from Ref [5].

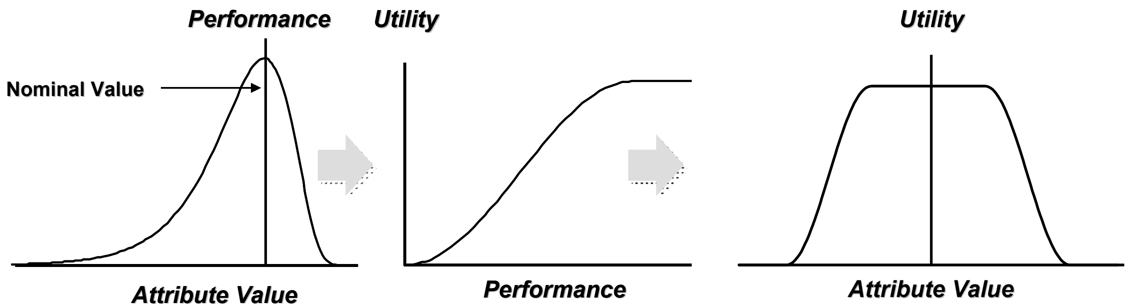

Figure 3. **End Item Performance and Attribute Value**. The performance of an end item parameter is a function of its associated measurable attribute. The utility of the parameter is, in turn, linked to its performance. Since performance is a function of attribute value, utility is also a function of end item attribute value.

The End Item Performance Model

At some time t elapsed since testing, the performance of the end item can be expressed as

$$P(\text{success}, t) = P_{sr} \int_{-\infty}^{\infty} f(x,t) P(\text{success} \mid x)\, dx \,, \tag{2}$$

where P_{sr} and $f(x,t)$ are described in Table 1. The variable P_{sr} is related to the usage context and may vary from application to application. The probability density function (pdf) $f(x,t)$ is a function of the distribution of values of the attribute of interest, emerging from the testing process, and of a time-dependent "uncertainty growth" mechanism. The construction of the pdf is described later under end item testing.

The "average" probability of successful performance of the supported end item function is obtained by integrating Eq. (2) over the end item's expected life cycle T

$$\begin{aligned} P(\text{success}) &= \frac{1}{T} \int_0^T P(\text{success}, t) dt \\ &= \frac{P_{sr}}{T} \int_0^T \int_{-\infty}^{\infty} f(x,t) P(\text{success} \mid x)\, dx\, dt \,. \end{aligned} \tag{3}$$

Note that, if the end item is periodically tested, T represents the interval between tests [1, 2].

END ITEM TESTING

As stated in the previous section, the pdf $f(x,t)$ is a function of the distribution of values of an end item attribute of interest, emerging from the testing process, and of a time-dependent "uncertainty growth" mechanism. The former is referred to as the "post-test distribution." The pdf for this distribution is developed below using the quantities defined in Table 2.

Table 2. End Item Testing Variables

Variable Description	Variable Name
Post-test probability density function	$f_{pt}(x)$
Pdf for attribute values not adjusted during testing	$f(x\|\text{not adjust})$
Probability that the attribute value will not be adjusted during testing	$P(\text{not adjust})$
Pdf for attribute values that are adjusted during testing	$f(x\|\text{adjust})$
Probability that the attribute value will be adjusted during testing	$P(\text{adjust})$
Symmetric ± adjustment or "guardband" limits	L_{adj}
Pdf for attribute values prior to testing	$f_{bt}(x)$
Pdf for measured end item values obtained during testing	$f_y(y\|x)$

The Post-Test Distribution

The post-test distribution pdf is the pdf $f(x,t)$ at time $t = 0$:

$$f_{pt}(x) = f(x,0) . \tag{4}$$

This function can be constructed from two additional pdfs and two probability functions

$$f_{pt}(x) = f(x \mid \text{not adjust})P(\text{not adjust}) + f(x \mid \text{adjust})P(\text{adjust}) , \tag{5}$$

where $f(x|\text{not adjust})$, $P(\text{not adjust})$, $f(x|\text{adjust})$ and $P(\text{adjust})$ are defined in Table 2. The first component of the right-hand side of Eq. (5) is obtained using the Bayes' relation described in standard statistical texts

$$f(x \mid \text{not adjust})P(\text{not adjust}) = f(\text{not adjust} \mid x) f_{bt}(x) \;\; , \tag{6}$$

and Eq. (5) becomes

$$f_{pt}(x) = f(\text{not adjust} \mid x) f_{bt}(x) + f(x \mid \text{adjust})P(\text{adjust}) . \tag{7}$$

If the variable y represents an observed (measured) attribute value whose true value is x, and $\pm L_{adj}$ represent adjustment or "guardband" limits, then the pdf $f(\text{not adjust}|x)$ is given by

$$f(\text{not adjust} \mid x) = \int_{-L_{adj}}^{L_{adj}} f_y(y \mid x) dy , \tag{8}$$

where $f_y(y|x)$ is the pdf for obtaining a value of y given the measurement of an end item attribute value x.[3] Reference [1] develops the above expressions for cases where all pdfs represent normally distributed variables. Cases where adjustments to center spec are made for every tested attribute (renew always) and cases where adjustments are made only for values lying outside guardband limits (renew as-needed) are both considered.

With this treatment, the pdf $f(x|\text{adjust})$ is constructed from the pdfs $f_{bt}(x)$ and $f_y(y|x)$. For example, if

$$f_{bt}(x) = \frac{1}{\sqrt{2\pi}\sigma_{bt}} e^{-x^2/2\sigma_{bt}^2} \;\; , \tag{9}$$

and

$$f_y(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_{tp}} e^{-(y-x)^2/2\sigma_{tp}^2} , \tag{10}$$

then

[3] Note that this expression applies to symmetric two-sided guardband limits. Extension to cases with asymmetric guardband limits or single-sided limits is straightforward.

$$f(x\,|\,\text{adjust}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-x^2/2\sigma_t^2} \, , \tag{11}$$

where

$$\sigma_t = \sqrt{\sigma_{bt}^2 + \sigma_{tp}^2} \, . \tag{12}$$

Note that σ_{bt} is the uncertainty in the bias of the end item attribute value prior to testing and σ_{tp} is the measurement uncertainty of the test process, including the uncertainty due to possible errors introduced by attribute value adjustment. The above relations correspond to post-test distributions such as in Figure 4.

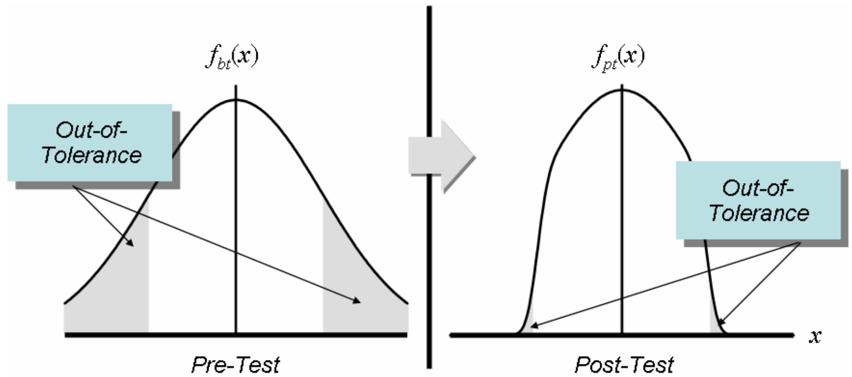

Figure 4. **A Typical Post-Test Distribution**. Shown is a case where the end item attribute value is normally distributed prior to testing. The distribution following testing improves the end item in-tolerance probability and changes the shape of the distribution. The pdf $f_{pt}(x)$ was computed using Eq. (7).

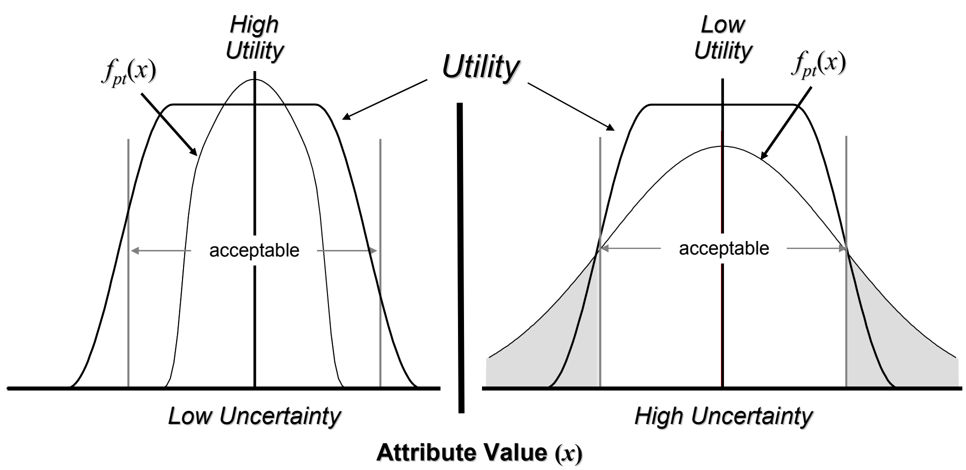

Figure 5. **Post-Test Distribution vs. Utility**. The probability of a particular utility of an end item function is given by the probability of the occurrence of a given attribute value. The figure on the left corresponds to high TS accuracy while the figure on the right corresponds to low TS accuracy. Clearly, lower TS accuracy (higher uncertainty) yields a lower probability for high utility.

The End Item Attribute In-Use Distribution

Returning to Eq. (4), we seek to develop a pdf $f(x,t)$ from the pdf $f_{pt}(x)$. For the case of normally distributed $f(x|\text{adjust})$ and $f(\text{not adjust}|x)$ pdfs, this can be accomplished simply by including time-dependent terms in the standard deviations of the pdfs. For example, Eq. (12) would become

$$\sigma_t = \sqrt{\sigma_{bt}^2 + \sigma_{tp}^2 + \sigma_u^2(t)} \,, \tag{13}$$

where $\sigma_u(t)$ provides for the growth of σ_t from the time elapsed since testing.

In most cases, stresses and stress responses during use are assumed to obey the following

1) Responses to stress of attributes toward increasing values occur with equal probability to responses toward decreasing values.
2) Attribute responses to stress are random in magnitude.
3) Stresses occur at some average rate r.
4) Stresses occur at intervals of some average duration of time τ.
5) Stresses are normally distributed with standard deviation σ_ζ.

Given these assumptions, responses to usage stress are seen to follow the classic random walk behavior. Letting the variable ζ represent the value of a change in end item attribute value due to a usage stress event, the pdf for stress response can be written

$$q(\varsigma \mid x) \equiv \frac{e^{-(\varsigma - x)^2/2\sigma_s^2}}{\sqrt{2\pi}\sigma_s} \,, \tag{14}$$

where $\sigma_s^2 = \sigma_\zeta^2 \dfrac{t}{\tau}$. Then the pdf $f(x,t)$ can be expressed as

$$f(x,t) = \int_{-\infty}^{\infty} f_{pt}(x) q(\zeta \mid x) d\zeta \,. \tag{15}$$

Ordinarily, stresses encountered during use are believed to "randomize" end item attribute bias distributions with the result that, over time, they approach the normal distribution.

In addition to uncertainty growth, the bias in an end item attribute may change systematically with respect to time. This effect is referred to as "drift." Attribute value drift is not currently included in the ETS model. This will be rectified through the incorporation of recent work [6]. This work will also be updated and included in an upcoming revision of NCSLI RP-1 [7].

PROBABILITY OF SUCCESSFUL END ITEM PERFORMANCE

The pdf $f(x,t)$ of Eq. (15) and the function $P(\text{success}|x)$ of Eq. (1) are employed to obtain the probability of successful performance $P(\text{success}, t)$ as a function of time elapsed since testing. We now equate the probability $P(\text{success}|x)$ with the end item attribute utility function in Eq. (1). This yields

$$\begin{aligned} P(\text{success}, t) &= P_{sr} \int_{-\infty}^{\infty} f(x,t) P(\text{success} \mid x)\, dx \\ &= P_{sr} \int_{-x_d}^{x_d} f(x,t) dx + 2P_{sr} \int_{x_d}^{x_f} \left\{ 1 - \sin^2 \left[\frac{(x - x_d)\pi}{2(x_f - x_d)} \right] \right\} f(x,t) dx \,. \end{aligned} \tag{16}$$

The "average" probability of successful performance $P(\text{success})$ of the supported end item function is obtained by integrating Eq. (16) over the end item's expected life cycle T, as shown in Eq. (3).

TEST AND CAL SYSTEM CALIBRATION

Test System Calibration

Each test system (TS) is assumed to be calibrated at periodic intervals. The start of each interval is termed the "beginning of period" (BOP), and the end of each interval is called the "end of period" (EOP). The beginning of period starts upon receipt of the TS by its user, and the end of period is marked at the point where the TS is sent for calibration by the user facility. The TS is calibrated periodically to control the measurement biases of its various attributes and the spreads in the attribute bias distributions. Each of these variables is considered to be a function of time elapsed since calibration. The spreads of the bias distributions are quantified by the distribution standard deviations which are synonymous with TS bias uncertainties.

As was shown earlier, the distribution f_{pt} of an end item attribute at the onset of usage is determined by the statistical bias distributions of the test systems and by the measurement uncertainties of the test processes. At the time of TS calibration, the TS bias uncertainties are referenced to their EOP values. Effects of shipping and storage are accounted for by convolving the post-calibration TS distribution with a stress response distribution such as is shown in Eq. (14).

When applied to end item testing, the TS may be selected at any point in its calibration interval. The bias uncertainty is taken to be referenced to the time elapsed since the TS was calibrated. In developing a cost-based picture, this uncertainty may be taken to be an "average-over-period" (AOP) value.

The TS calibration process is characterized by several sources of uncertainty, which are quantified by the following set of standard deviations:

σ_{EOP} = the standard deviation of TS attribute values following shipping to the calibration facility (i.e., at EOP).

σ_s = the standard deviation of the response of TS attribute values to stresses encountered during shipping (set to zero if the TS is not shipped from the calibration facility).

σ_{cs} = the standard deviation of the calibration system (CS) attribute values at time of test or calibration. This is referred to as the bias uncertainty in the calibration reference. If random demand of CS items over the CS calibration interval is assumed, this is set equal to the AOP value.

σ_{cp} = the standard deviation of the calibration process error distribution. Equated with the calibration process uncertainty (excluding σ_{cs}).

These variables are employed in Refs [1] and [2] to provide a description of TS bias distributions following calibration. Provision for growth of the uncertainty σ_{cs} over the CS calibration interval is also addressed. In the ETS model, this uncertainty growth is estimated using the time-dependence of in-tolerance probabilities or "measurement reliabilities."

For purposes of discussion in this paper, it will suffice to say that the determination of post-calibration TS attribute biases is essentially the same as the determination of post-test end item attribute biases. The exceptions are the means of accounting for the time-dependence of TS attribute bias uncertainty and the fact that explicit considerations of TS utility are not relevant. This is because the utility of test and also of calibration systems is reflected in their support of the utility of the end items.

Calibration System Calibration

In the ETS model, the calibration of the calibration system (CS) is treated in the same manner as the calibration of the test system with "TS" replaced by "CS" and the standard deviation σ_{cs} referring to the uncertainty in the value of the attribute of the calibration standard.

CONTROLLING MEASUREMENT DECISION RISK

In Eq. (7) the function P(adjust) appears as key to the computation of successful end item performance. The conditional adjustment probability function is given by

$$P(\text{adjust}\,|\,x) = \begin{cases} 1, & x < A_1 \text{ or } x > A_2 \\ 0, & \text{otherwise} \end{cases}, \tag{17}$$

where A_1 and A_2 are defined in Table 3. From this, the function P(adjust) is readily obtained using the pdf $f_{bt}(x)$

$$\begin{aligned} P(\text{adjust}) &= \int_{-\infty}^{\infty} P(\text{adjust}\,|\,x) f_{bt}(x)dx \\ &= 1 - \int_{-A_1}^{A_2} f_{bt}(x)dx\,. \end{aligned} \tag{18}$$

A_1 and A_2 can be adjusted relative to the attribute tolerance limits L_1 and L_2 to control measurement decision risks associated with the calibration or test process.[4]

Table 3. Risk Control Variables

Variable Description	Variable Name
UUT attribute value	x
Lower and upper tolerance limits for x	L_1 and L_2
Lower and upper acceptance (guardband) limits for x	A_1 and A_2
Measured value obtained for x during testing or calibration	y
Joint probability of events E1 and E2 happening	P(E1,E2)
Conditional probability that event E2 will happen if event E1 has occurred	P(E2|E1)

The measurement decision risk functions commonly of interest are *false accept risk* (FAR) and *false reject risk* (*FRR*). *FAR* comes in two varieties. One is of interest to the testing or calibration facility. This is called *unconditional FAR* or *UFAR*. The other is referenced to the equipment user. This is called *conditional FAR* or *CFAR* [8].

In most applications, the *UFAR* variety is used. In much of the statistical literature, *UFAR* is also referred to as *consumer's risk* [9, 10]. False reject risk is sometimes called *producer's risk*. These risks are defined as

- *UFAR* - probability that the UUT is out-of-tolerance and observed to be in-tolerance during testing or calibration
- *CFAR* - probability that the UUT is out-of-tolerance given that it was accepted as being in-tolerance
- *FRR* - probability that the UUT is in-tolerance and observed to be out-of-tolerance during testing or calibration

Letting A represent the range A_1 to A_2 and L represent the range L_1 to L_2, these risks can be expressed in terms of probability functions [3,8]. Using the notation of Table 3, we have

[4] Note that, for single-sided tolerances, L_1 or L_2 may be set to $-\infty$ or $+\infty$, respectively. Note also that for renew always policies, $A_1 = A_2 = 0$, so that P(adjust) = 1.

$$
\begin{aligned}
UFAR &= P(x \notin L, y \in A) \\
&= P(y \in A) - P(x \in L, y \in A) \\
CFAR &= P(x \notin L \mid y \in A) \\
&= \frac{UFAR}{P(y \in A)} \\
FRR &= P(x \in L, y \notin A) \\
&= P(x \in L) - P(x \in L, y \in A).
\end{aligned}
$$

These probabilities can also be written

$$
\begin{aligned}
UFAR &= P(A_1 \le y \le A_2) - P(L_1 \le x \le L_2, A_1 \le y \le A_2) \\
CFAR &= \frac{UFAR}{P(A_1 \le y \le A_2)} \\
FRR &= P(L_1 \le x \le L_2) - P(L_1 \le x \le L_2, A_1 \le y \le A_2).
\end{aligned}
$$

It can be readily appreciated that smaller values of A_1 and A_2, relative to L_1 and L_2, result in reduced *UFAR* and *CFAR*. Unfortunately, they also result in increased *FRR*.

While controlling risks can alter post-test and post-calibration distributions in favorable ways, setting risk control criteria is not especially productive from a cost perspective. This is because the means for obtaining explicit estimates for the cost of a false accept and the cost of a false reject are not readily available. In addition, these costs are not easily incorporated in an integrated cost model.

The ETS model does not apply specific decision risk criteria to optimize costs. However, the ETS model *does* control risks to optimize costs. This is done partly through adjusting the limits A_1 and A_2. Adjusting these limits is one way to achieve cost optimization. Other risk control variables are aimed at controlling TS and CS attribute bias uncertainty. These include calibration interval adjustment and TS and CS equipment selection.

COST OPTIMIZATION

Total Cost

The ETS model leads to cost optimization through an estimate of the total cost C_{tot}

$$C_{tot} = C_{acc} + C_{ts} \quad . \tag{19}$$

The quantity C_{acc} is the end item "acceptance cost," expressed as a function of the cost of reduced utility in the context of various usage scenarios, each weighted by a probability of occurrence. The quantity C_{ts} is the total annual test and calibration support cost. This is a function of the variables shown in Tables 4 and 5 and of various probabilities, such as *P*(adjust), *P*(success) and so on [1].

Cost Modeling

Calibration intervals, test decision risks, and availability are parameters which have a direct bearing on the costs associated with operating and maintaining a test and calibration support hierarchy. These parameters also impact indirect costs associated with end item quality and/or performance capability.

Table 4. ETS Cost Modeling Variables

Variable Description	Variable Name
End item attribute value corresponding to the onset of degraded performance	x_d
End item attribute value corresponding to loss of function	x_f
Cost of a given outcome	C_f
Quantity of end items in use or in inventory	N_{UUT}
Acquisition cost of an end item unit	C_{UUT}
End item spare coverage desired (in percent)[5]	S_{UUT}
Probability of a successful outcome, given successful end item performance	P_{sr}
Probability of an encounter	P_e
Hours to calibrate/test	H_c
Additional hours required for adjustments	H_a
Cost per hour for test/calibration and/or adjustment	C_{hr}
Cost per repair action	C_r

End item performance capability is measured in terms of the extent to which an end item achieves a desired effect or avoids and undesired effect. These effects can be referenced to the program management level, in the case of a military or similar system, to an end item producer in the case of a commercial product or to any category of end item disposition for which the end item quality and/or performance capability can be quantified in economic terms. Examples of desired effects may include a successful strike of an offensive weapon, follow-on reorders of a product item, creation of a desirable corporate image, etc. Examples of undesired effects may include an unsuccessful response to a military threat, warranty expenses associated with poor product performance, return of products rejected by customer receiving inspection, etc. In each case, the end item experiences an "encounter" (approach of an intended target, approach of an incoming missile, appearance of an unexpected highway obstruction, etc.) which results in a perceived "outcome" (successful missile strike, missile interception, obstruction avoidance, etc.). The effect is determined by the "response" of the end item to the encounter (timely sighting and ranging, early detection and warning, nominal braking and maneuvering, etc.). The cost of a given outcome is a variable which is assumed to be known. If an outcome is associated with a benefit, the cost is expressed as a negative quantity.

Table 5. ETS Acceptance Cost Modeling Variables

Variable Description	Variable Name
Total annual cost	C_{tot}
Annual acceptance cost	C_{acc}
Total annual support cost	C_{ts}
Annual calibration cost	C_{cal}
Annual adjustment cost	C_{adj}
Annual repair cost	C_{rep}
Total spares acquisition cost	C_{sa}

The variables employed in modeling acceptance cost are shown in Tables 4 and 5. In Table 5, total annual calibration, adjustment, repair, and support costs relate to costs incurred from support of an item of interest (calibration system, test system or end item). Annual acceptance cost applies only if the item of interest is an end item.

The acceptance cost for end items is the product of the cost of a given outcome, the number of end items in use, the probability of an encounter occurring and the probability of unsuccessful end item performance:

[5] This variable controls the number of spares maintained to cover the end item inventory or the population of end items sold to customers.

$$C_{acc} = C_f N_{UUT} P_e [1 - P(\text{success})] \ , \tag{20}$$

where P(success) is given in Eq. (3).

The quantity C_{acc} can be "annualized" by expressing P_e in terms of the probability of encounter per end item unit per year. As stated earlier, acceptance cost applies only to the end item.

Other cost variables contributing to the total support cost C_{ts} are

$$C_{ts} = C_s^{year} + C_{cal} + C_{adj} + C_{rep} \ . \tag{21}$$

where C_s, C_{cal}, C_{adj} and C_{rep} are defined in Table 5. The details of their estimation are given in Refs [1] and [2].

Accounting for Life Cycle Costs

The ETS model provides a rudimentary life cycle cost model that accounts for costs involved in equipment acquisition, spares requirements, test and calibration and equipment repair. A number of equipment life cycle cost variables, such as training costs, disposal costs and costs associated with resale need to be included for completeness. Work has been initiated toward this end [6].

Data Continuity Requirements

For the ETS model to work, an unbroken thread of data is required between all relevant levels of the test and calibration support hierarchy. In this, end item test plans need to focus on testing attributes relating to end item performance, and TS calibration procedures need to provide data on TS attributes that support end item performance. From experience in using ETS, the interface between test plans and end item performance requirements has often been virtually nonexistent [5]. Of course, in addition to facilitating use of the ETS model, this situation needs to be rectified to allow validation of end item test and calibration support in general. Efforts are underway to alleviate the problem [11].

APPLICABILITY OF NCSLI RECOMMENDED PRACTICES

Many of the variables and relationships of the ETS model are covered by existing NCSLI recommended practices and by others under development. The relevant RPs are listed in Table 6.

Table 6. NCSLI RPs Relevant to the ETS Model

Controlling uncertainty growth in attribute bias	RP-1.
Obtaining, evaluating, developing and using equipment specifications	RP-5.
Estimating and controlling measurement uncertainty	RP-12
Estimating and controlling measurement decision risks	Under development.
Optimizing program decisions in managing the cal/test hierarchy	Under development.
Controlling measurement processes with SPC	Under development.

REFERENCES

[1] Castrup, H., "Calibration Requirements Analysis System," *Proc. NCSL Workshop & Symposium*, Denver, 1989.

[2] NASA, *Metrology — Calibration and Measurement Processes Guidelines*, NASA Reference Publication 1342, June 1994.

[3] Integrated Sciences Group, *Analytical Metrology Handbook*, www.isgmax.com, 2006.

[4] NASA, *Systems Engineering Handbook*, NASA/SP-2007-6105 Rev1, Under Development.

[5] Castrup, H., "Quantifying Test Quality Improvement," Navy Metrology R&D Requirements Conference, Annapolis, June 1993.

[6] Castrup, H., "Procurement Analysis Feasibility Study," Slide Presentation, Unpublished, March 2000.

[7] NCSLI, *Establishment and Adjustment of Calibration Intervals*, Recommended Practice RP-1, National Conference of Standards Laboratories, January 1996.

[8] Castrup, H., " Risk Analysis Methods for Complying with Z540.3," *Proc. NCSLI Workshop & Symposium*, St. Paul, 2007.

[9] Eagle, A., "A Method for Handling Errors in Testing and Measuring," *Industrial Quality Control*, March, 1954.

[10] Grubbs, F. and Coon, H., "On Setting Test Limits Relative to Specification Limits," *Industrial Quality Control*, March, 1954.

[11] NASA, *Measurement Quality Assurance*, Handbook, Under Development.