General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

Measurement Decision Risk – The Importance of Definitions

Speaker/Author:
Scott M. Mimbs
National Aeronautics and Space Administration (NASA)
John F. Kennedy Space Center
Kennedy Space Center, FL 32899
Phone: 321-861-5184
Fax: 321-867-1740
scott.m.mimbs@nasa.gov

Abstract

One of the more misunderstood areas of metrology is the Test Uncertainty Ratio (TUR) and its cousin, the Test Accuracy Ratio (TAR). There have been many definitions over the years, but why are these definitions important to a discussion on measurement decision risk? The importance lies in the clarity of communication. Problems can immediately arise in the application (or misapplication) of the definition of these terms. In other words, while it is important to understand the definitions, it is more important to understand concepts behind the definitions and to be precise in how they are applied.

The objective of any measurement is a decision. Measurement Decision Risk is a way to look at the quality of a measurement and although it is not a new concept, it has generated a lot of attention since its addition as a requirement in the new U.S. National Standard, ANSI/NCSL Z540.3-2006. In addition to Measurement Decision Risk as the prime method of managing measurement risk, Z540.3 has also added, as a fall-back, an explicit definition for TUR. The impact these changes might have on calibration service providers if these requirements are levied on them has become the topic of much discussion and in some cases concern.

This paper looks at the concepts behind the definitions and how they relate to Measurement Decision Risk. Using common examples, this paper will also provide a comparison of various elements of risk related to measurement science using the concepts of TAR, TUR, accuracy ratios, and Consumer Risk (False Accept Risk). The goal is to provide a better understanding of their relevance to the measurement decision process.

Introduction

The international definition of metrology is the science of measurement and measurement is further defined as the set of operations having the object of determining the value of a quantity [3]. Although metrology is about measurements, the real bottom-line is what is done with the information from the measurements. In the simplest terms, measurements are made to support decisions and/or establish facts. For example, measurements help make decisions:

- To continue or stop a process (including a space launch)
- To accept or reject a product
- To rework or complete a design

- To take corrective action or withhold it
- To establish scientific or legal fact.

If the data from measurements are not being used in a decision or to establish facts (including scientific research), the measurement is unnecessary. *"The more critical the decision, the more critical the data. The more critical the data, the more critical the measurement"* [1].

Assuring the validity of the measuring and test equipment (M&TE) is an essential component of the measurement process. Calibration has been a part of the measurement process since the Egyptian Cubit Stick was used in the building of the pyramids [2]. As technology increased in the mid-20th century, the need for more precision in testing and calibration led to the development of new methods in determining the risks involved in measurement related decisions. The TUR and TAR developed out of these activities and have been used for over 50 years as tools for mitigating measurement decision risk. But what was lacking with the TUR/TAR was a general consensus on how to apply them or even how to define them.

As with any new consensus standard, organizations owning and using M&TE will have to review the Z540.3 for potential impacts to their business (positive and negative) and make the decision whether or not to incorporate the standard. Organizations and calibration service providers may have to bid on contracts that include the standard and will need to understand the implications of measurement decision risk and the explicit definition of TUR. This paper attempts to address these concepts by looking at the following areas.

1. A look at the history and development of measurement decision risk and the TUR.
2. Discuss the current definitions of TUR/TAR.
3. Discuss how the Z540.3 definition of TUR is linked to measurement decision risk and the limitations of its use.
4. Review an example of the application of a TUR in an off-nominal case and what the measurement decision risk calculates to be for such usage.

Development of Measurement Decision Risk

A look at the origin of the TUR/TAR helps in understanding their relationship to measurement decision risk.

Measurement decision risk analysis traces its roots to the early work on consumer and producer risk analysis done by Alan Eagle, Frank Grubbs, and Helen Coon [5,6] in the late 1940's and early 1950's. Eagle's 1954 paper describes the methods for calculating the consumer and producer risk and how to establish "test limits" in relation to design limits which have become known as guardbands today. The focus of the paper was to analyze and mitigate the "test errors" which are "inherent in the test equipment and/or test personnel" used in the inspection of manufactured complex electronic equipment [5]. The key point to Eagle's method was quantifying and using consumer/producer risk (measurement decision risk) as a part of the manufacturing process. This concept is applicable to any application where decisions are based on measurements.

In 1955, the U.S. Navy recognized the need for better measurement reliability in their guided missile program. Jerry Hayes of the U.S. Naval Ordnance Laboratory authored a Technical Memorandum (TM) which built upon Eagle's work [7]. Here, Hayes set out to establish a basis for accuracy ratios versus decision risks for application in the Navy's calibration program. The practice at the time was to use a 10:1 ratio, but that value was considered unsupportable by the nation's calibration support and measurement traceability infrastructure. Using the relationship between the design specifications, testing limits, and instrument error, Hayes proposed using a "family of curves" to determine the specific testing risk or reliability. The problem with this method was a new family of curves had to be established each time a process or design tolerance changed. A change in a process or tolerance nullified the assumptions upon which the first set of curves was built [7, 8]. Some important aspects of the Hayes paper, still relevant today, are the need for calibrated equipment used in testing, establishment of reasonable testing risk levels, reasonable design tolerances, and adequate procedures for testing.

After the release of the 1955 TM, Hayes continued to work on methods of assuring measurement reliability based on consumer risk. In the mid-50's, computing consumer risk was a very arduous task (requiring use of a slide rule), which Hayes decided not to require U.S. Navy contractors to perform. With very specific assumptions on process, a consumer risk of 1% was selected, which calculated to be about a 3:1 accuracy ratio. Hayes, working with Stan Crandon, decided to pad the ratio to account for uncertainty in the reliability of the tolerances industry was using for the measurement standards (at the time, only tolerances were used in both the numerator and denominator when figuring accuracy ratios). Thus the 4:1 ratio requirement was developed and established as Navy policy and subsequently adopted as a requirement in military procurement standards both here and abroad [8]. This ratio became what is known today as the TAR and later evolved into the TUR.

A Look at Existing Definitions

Various definitions have appeared in many texts, papers and documents over the years, but there was not a consensus standard available that provided a definition or specified how to apply the rule. Over time, two distinct terms surfaced – Test Accuracy Ratio (TAR) and Test Uncertainty Ratio (TUR). The following are some current definitions.

- The American Society for Quality defines TAR and TUR in terms of calibration [2].

  Test accuracy ratio - (1) In a calibration procedure, the test accuracy ratio (TAR) is the ratio of the accuracy tolerance of the unit under calibration to the accuracy tolerance of the calibration standard used. [...]

  $$TAR = \frac{UUT\_Tolerance}{Std\_Tolerance}$$

  Test uncertainty ratio - In a calibration procedure, the test uncertainty ratio (TUR) is the ratio of the accuracy tolerance of the unit under calibration to the uncertainty of the calibration standard used. [...]

  $$TUR = \frac{UUT\_Tolerance}{Std\_Uncertainty}$$

- Although a direct definition of TUR is avoided, NASA's Space Shuttle Program has ratio requirements that apply to calibration and article or material measurements [9].

Paragraph 4: Article or Material Measurement Processes
The Expanded Uncertainty in any article or material measurement process shall not exceed ten percent of the tolerance of the article or material characteristic being measured. [...]

Paragraph 5: Calibration Measurement Processes
[...] the Expanded Uncertainty in any calibration measurement process shall not exceed 25 percent of the tolerance of the parameter being measured. [...]

- There have been times when the two terms (TAR & TUR) were considered interchangeable. This was documented in the Z540.1 Handbook [10].

Interpretive Guidance for Section 10.2 of the Handbook [10]

As a default alternative to doing an uncertainty analysis, a laboratory may rely on a Test Accuracy Ratio (TAR) of 4:1. A TAR of 4:1 means that the tolerance of the parameter (specification) being tested is equal to or greater than four times the combination of the uncertainties of all the measurement standards employed in the test.

If it is determined that the TAR is less than 4:1, then one of the following methods may be used: uncertainty analysis as described above, guard-banding, widening the specification, or another appropriate method.

Note: Some refer to TARs as Test Uncertainty Ratios or TURs

- The Z540.3 provides an explicit definition of TUR, but does not address the TAR.

3.11 Test uncertainty ratio [4]
The ratio of the span of the tolerance of a measurement quantity subject to calibration, to twice the 95% expanded uncertainty of the measurement process used for calibration.

NOTE: This applies to two-sided tolerances. [...]

The definition uses the expanded uncertainty as defined in ANSI/NCSL Z540-2-1997[NCSL, 1997] where (k) is the coverage or confidence factor. The definition in equation form:

$$\mathrm{TUR} = \frac{\mathrm{Upper} - \mathrm{Lower}}{2 \cdot U} \qquad U = k \cdot u \qquad k = 2$$

There are key differences between the Z540.3 and earlier TUR definitions.

1. The earlier TUR/TAR denominator is not well defined which leads to inconsistent applications.
2. The denominator for the Z540.3 TUR is explicitly defined, including the confidence level for the coverage factor (k), thus providing better uniformity in the application of the TUR.

Although it only appears in the ASQ definitions, the TAR is one of the more popular applications of the risk rules-of-thumb. It is applied in many labs and other applications and can be found in many papers and training guides for calibration and quality inspection. It must be cautioned however, as such, the TAR is non-compliant with either Z540.1 or Z540.3.

As mentioned earlier, the objective of a measurement is a decision and Measurement Decision Risk is a tool to assess the suitability of the measurement process. Even in their different forms, TARs and TURs are used to mitigate the risk in measurement. The value of a concise definition is important for clarity of communication - problems can quickly arise when different meanings are applied to the same term. The devil is in the details.

Linking the TUR to Measurement Decision Risk

It is from Eagle's work the TUR was first developed [7, 8]. For this reason the consumer risk equation associated with Eagle's work [5, 6] is presented here again and discussed in detail to help in explaining the relationship of the TUR to Measurement Decision Risk. Although there are other variations of measurement decision risk equations, they will not be discussed in this paper.

$$CR = \frac{1}{\pi} \cdot \int_{k}^{\infty} \int_{-r \cdot (k+t)+b}^{r \cdot (k-t)-b} e^{\frac{-(t^2+s^2)}{2}} \, ds \, dt$$

This equation has three distinct variables, r, k, and b which influence the results.

The variables will be considered in reverse order. The variable (b) is the deviation of the test limits from the specification limits [i.e., $\mu \pm (k\sigma_X - b\sigma_e)$]. This establishes the required guardband (or "test limit") to achieve a desired consumer risk. For the purposes of this discussion, the specification limits will equal the test limits (i.e., b = 0), therefore leaving only two influence variables for this discussion.

The variable (k) is the number of standard deviations the performance specification limit is to the product mean, which is assumed to be centered (i.e., $\mu \pm k\sigma_X$). If the measured value lies between $\pm k\sigma_X$, then the subject parameter is conforming. Figure 1 illustrates the relationship of the specification limits and production (or process) distribution with the limits falling at $\pm 2\sigma$

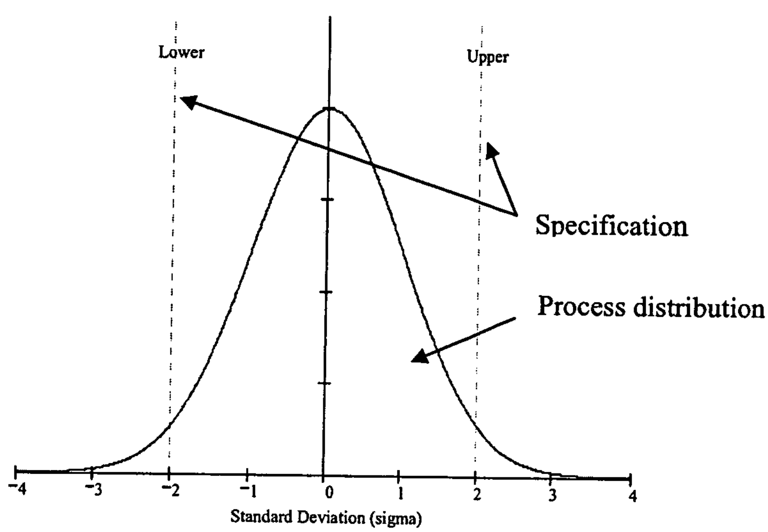

Figure 1: The relationship of the specification limits to the process distribution with the limits falling at ±2σ.

The last variable to consider is (r) which is a ratio of two standard deviations.

$$r = \frac{\sigma_X}{\sigma_e}$$

The numerator σ_X is what Eagle described as "the true standard deviation of the product distribution." This represents the subject of interest. Eagle described the denominator σ_e as "the standard deviation of the errors of measurement." This represents the measurement process with the mean assumed to be zero [5]. In metrology, the standard deviation is an indicator of uncertainty (i.e., $\pm$ kσ indicates $\pm$ ku), thus the variable (r) may be considered an "accuracy ratio" (of sorts).

$$r = \frac{\sigma_X}{\sigma_e} = \frac{u_X}{u_e}$$

Eagle noted in his original paper that if the "performance specification" falls at more than 2 standard deviations of the production distribution (i.e., k > 2), then mitigation efforts are not required [5]. His statement may not be true for all cases in today's technology, but this was the assumption that Hayes and Crandon used to establish the 4:1 rule in the mid-1950s [8]. Figure 2 shows the "knee" of the curve intersects the 1% risk line at approximately an r-value of 3:1. As 1% was the risk level Hayes was seeking, with a little "padding," 4:1 became the "rule of thumb" assuming the production (or process) distribution was equal to or better than $2\sigma_X$.

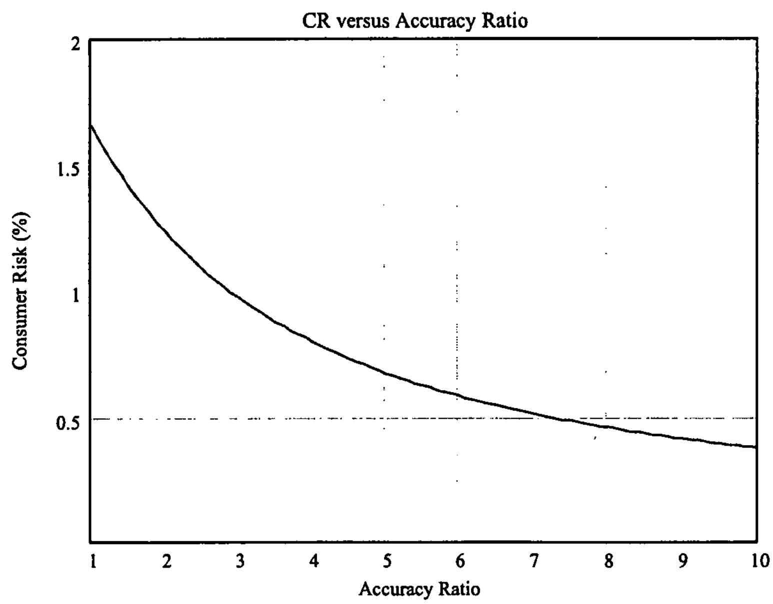

Figure 2: Consumer Risk versus the "accuracy ratio" for the specific case where the specification limits lie at 2 standard deviations from the mean of the production distribution (i.e., k=2 for $\mu \pm 2\sigma_X$).

Although this "accuracy ratio" is considered the origin of the TUR, it is quite different, the most noteworthy being there are no confidence limits (±k) to either the numerator or denominator. The ratio is a pure "uncertainty ratio."

The amount of work required in 1955 to perform the calculations was extraordinary using slide rules and paper. Hayes and Crandon used a chart similar to Figure 2 for their analysis as they searched for a method to improve measurements without levying the large amount of calculations on the contractors. When Hayes allowed the use of a ratio between the tolerances of the subject of interest and the measuring equipment, the idea was supposed to be temporary until a method could be developed using a more sound method and better computing power [8]. As discussed earlier, Hayes settled on 4:1, but others went with a more conventional and conservative 10:1. NASA used the 10:1 for all calibration and article measurement requirements through the first moon landing in 1969. After that, calibrations were allowed to use 4:1 while test measurements remained at 10:1 [9].

As time has gone by, many have tried to help the TUR/TAR become a tool that is more soundly founded in measurement science. The Z540.3 comes very close to accomplishing that goal. Let's take another look at the ratio from the consumer risk equation, but now only extend the standard deviation to measurement process uncertainty.

$$r = \frac{\sigma_X}{\sigma_e} = \frac{\sigma_X}{u_e}$$

Remember that the relationship of the tolerance to the product mean is $\mu \pm k\sigma_X$. Combined with the value of (r) as stated above, the Z540.3 TUR can be transformed (slightly).

$$TUR = \frac{\text{Upper} - \text{Lower}}{2 \cdot U_{95}} = \frac{\text{Upper} - \text{Lower}}{2 \cdot k_e \cdot u_e} = \frac{2 \cdot k_X \cdot \sigma_X}{2 \cdot k_e \cdot u_e} = \frac{k_X \cdot \sigma_X}{k_e \cdot u_e}$$

As shown here, the TUR can be represented as a ratio of intervals (± kσ). For Z540.3, the coverage factor in the denominator is usually inferred to be k = 2 (95% of normal). But there is no inferable definition for the value of (k) in the numerator, although it might be assumed that the coverage factor in the numerator represents the (k) factor from the consumer risk equation. Knowing this information would complete the link of the TUR to measurement decision risk.

It has been suggested that for calibration, the End of Period Reliability (EOPR) for the Unit under Test (UUT) represent the "product distribution" [13]. EOPR is the probability of a unit being in-tolerance when it is returned for routine calibration at the end of its normal interval, thus EOPR fits the relationship of process versus limits as previously shown in Figure 1.

Using EOPR to represent the UUT product distribution opens up many questions about the level of "drill down" required in obtaining this information (e.g., nomenclature, manufacture, model number, serial number, or parameter value). The answers to these questions are beyond the scope of this paper, but an excellent detailed discussion can be found in Section 6.3 of NASA Reference Publication 1342 [1].

Applying Definitions – Example 1

A generic example of a UUT and a working standard will be used to illustrate the concepts just discussed. The uncertainty is a generic estimate with no assumptions defined, as it only represents the combined standard uncertainty of the measurement process

Parameter	UUT	Working Standard
Tolerance	±4.0	±1.0
$u_c = 0.67$		
TAR = 4:1		
TUR = 3:1		

Table 1: A generic calibration example.

With a TUR of less than 4:1, if the Z540.3 is applicable, this calibration would be unacceptable for application of the TUR fall-back rule. But is the calibration "good enough" – is the Measurement Decision Risk acceptable (2% or less per the Z540.3)?

As discussed above, the assumption in the creation of the original 4:1 ratio, with respect to the desired level of measurement risk, was 1% [8]. The value of (k) was a pivotal assumption because when k = 2 (in a normal distribution), 95% of the process would be contained within the specification limits. As stated earlier, the product distribution is to be represented by the EOPR of the UUT. Figure 3 illustrates the relationship between consumer risk and EOPR for TUR ratios of 4:1 and 3:1.

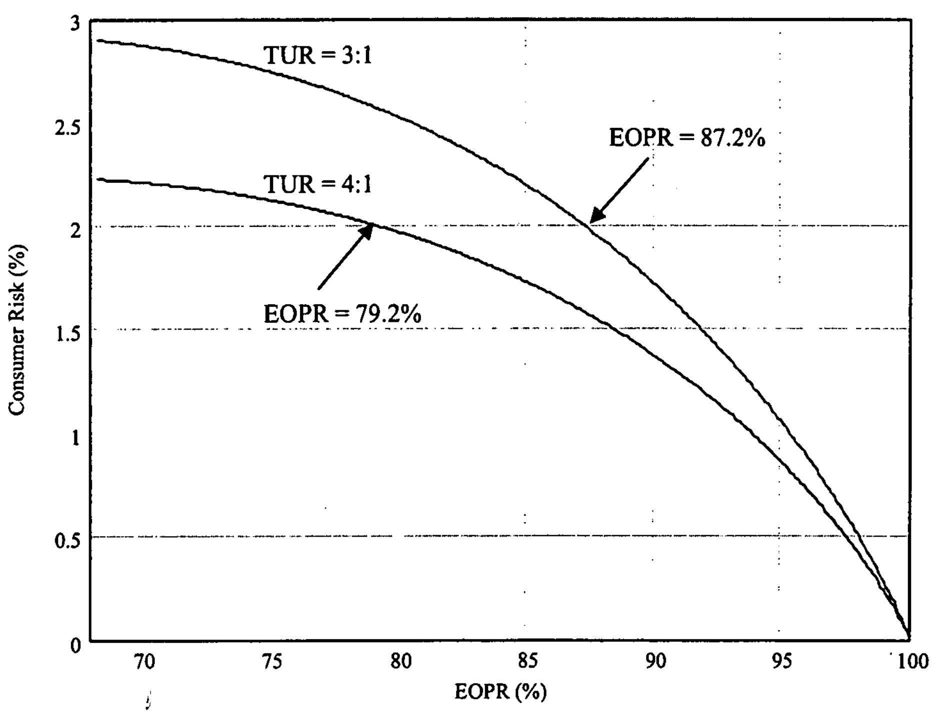

Figure 3: Consumer Risk versus the EOPR.

Figure 3 shows it is possible for a TUR below 4:1 to be acceptable (given 2% is acceptable), but it also demonstrates that 4:1 may not provide a desired level of risk. For example, as Figure 3 shows, the EOPR must be above 79% for a 4:1 to provide a consumer risk of 2%. For some organizations and applications, an EOPR of 79% may be unacceptable. The bottom-line is a TUR may not provide a proper or desired level of quality without additional information. Using the TUR alone may lull an organization into a false sense of security.

Example 2

All types of calibrations will be affected by Z540.3, but those that do not fit the general mold of the TUR or 2% rule are the calibrations that will create the most problems for quality mangers and calibration service providers. This example will look at the calibration of a digital micrometer, an operation performed millions of times a year across our planet. Table 2 contains the specification information for the micrometer and calibration standard (Class 2 Gage Blocks).

Parameter	**Digital Micrometer 0-1 inch (0-25.4 mm)**	**Gage Block, Class 2**
Tolerance	± 0.0001 inch (± 2.54 μm)	± 4.0 micro-inches (± 0.1016 μm)
Resolution	± 0.00005 inch (± 1.27 μm)	N/A

Table 2: The specification information for the digital micrometer and the calibration standard.

A new or overhauled micrometer is calibrated over at least 10 points of its range, but only 4 points for routine calibrations (0 plus low, middle, and high). The purpose of this example is to examine the measurement risk using TAR and TUR, therefore only one point of the calibration will be discussed - the mid point of 0.50000 inches (12.700 mm).

Using the ASQ definition, the calibration TAR has a value of 25:1.

$$\mathrm{TAR} = \frac{\mathrm{UUT\_tolerance}}{\mathrm{Std\_tolerance}} = \frac{0.0001}{0.000004} = 25$$

As previously stated, the purpose of this example (and paper) is the examination of usage of measurement risk tools. As such, the uncertainty analysis is presented in table form without the detailed calculations. The analysis follows the guidelines of the ANSI/NCSL Z540-2-1997, which is the U.S. equivalent of the ISO GUM [12]. The specific methods followed were those recommended in *Uncertainty Analysis Principles and Methods* [14], developed for the Naval Air Warfare Center Aircraft Division (NAWCAD) and Air Force Flight Test Center (AFFTC). Readers desiring the detailed calculations may contact the author. Table 3 contains the analysis results for the digital micrometer.

Uncertainty Source	Standard Uncertainty inches (μm)	Confidence Level (%)	Type (A or B)	Distribution
Gage Blocks	0.000002 (0.0508)	95.00	B	Normal
Resolution	0.0000144 (0.3658)	100.00	B	Uniform
Environmental	0.0000001 (0.00254)	95.00	B	Normal
Random Error or Repeatability	0.0000083 (0.2108)	95.00	A	Student's t
Combined Uncertainty	0.0000168 (0.4267)			

Table 3: Results of the uncertainty analysis for a digital micrometer.

The combined standard uncertainty from Table 3 is calculated into the Z540.3's TUR for the following result. For the 95% expanded uncertainty, assume k = 2.

$$\mathrm{TUR} = \frac{\text{Upper} - \text{Lower}}{2 \cdot U_{95}} = \frac{0.0002}{2 \cdot 2 \cdot 0.0000168} = 3$$

As can be seen, even with a very high TAR, when the calibration standard is passive (i.e., the UUT is the "reading" instrument), the TUR may fall below 4:1, in this case due to instrument resolution.

Example 3

The subject of Example 2 was a digital micrometer with a resolution that is one half of the unit tolerance. Many times the micrometer resolution and tolerance are equal, as in the case for Example 3. In this example, the micrometer has a vernier scale in addition to the regular graduations, allowing the user to read more precisely. The TAR for this example is 25:1, the same as for Example 2. Table 4 contains the specification information for this example.

Parameter	Analog (Vernier) Micrometer 0-1 inch (0-25.4 mm)	Gage Block, Class 2
Tolerance	± 0.0001 inch (± 2.54 μm)	± 4.0 micro-inches (± 0.1016 μm)
Resolution	± 0.0001 inch (± 2.54 μm)	N/A

Table 4: The specification information for the analog micrometer and the calibration standard.

Uncertainty Source	Standard Uncertainty inches (μm)	Confidence Level (%)	Type (A or B)	Distribution
Gage Blocks	0.000002 (0.0508)	95.00	B	Normal
Resolution	0.0000255 (0.6477)	95.00	B	Normal
Environmental	0.0000001 (0.00254)	95.00	B	Normal
Random Error or Repeatability	0.0000201 (0.5105)	95.00	A	Student's t
Combined Uncertainty	0.0000326 (0.8280)			

Table 5: Results of the uncertainty analysis for the analog micrometer.

The results of the uncertainty analysis are taken from Table 5 and again calculated into the Z540.3's TUR for the following result.

$$\mathrm{TUR} = \frac{\mathrm{Upper} - \mathrm{Lower}}{2 \cdot U_{95}} = \frac{0.0002}{2 \cdot 2 \cdot 0.0000326} = 1.5$$

As expected, the results are well below the 4:1 TUR required by both the Z540.1 and Z540.3.

Discussion of Example Results

The examples above were chosen for a reason – they do not meet the requirements of the existing or new standards for calibration in terms of the TUR. But do they meet the Z540.3's 2% rule for Measurement Decision Risk? The best way to answer that question is to calculate the consumer risk for the calibration. For illustrative purposes, the consumer risk for only the analog micrometer will be calculated.

The consumer risk equation will be set as a function of the variables (k), (r), and (b).

$$CR(k,r,b) = \frac{1}{\pi} \cdot \int_{k}^{\infty} \int_{-r\cdot(k+t)+b}^{r\cdot(k-t)-b} e^{\frac{-\left(t^2+s^2\right)}{2}} \, ds \, dt \qquad r = \frac{\sigma_x}{u_e}$$

At one of Kennedy Space Center's calibration labs, the nomenclature "micrometer" has an EOPR of just over 96%, but the model number of the analog micrometer has an EOPR of 97.3%. For an assumed normal distribution, this equals a coverage factor of 2.21 (i.e., k = 2.21).

$$\sigma_x = \frac{0.0001}{2.21} = 0.0000452 \qquad r = \frac{0.0000452}{0.0000326} = 1.387$$

Consumer Risk: $CR(2.21, 1.387, 0) = 0.9\%$

If the model specific EOPR information were not available, then the overall nomenclature EOPR of 96% would have yielded a consumer risk on the calibration of approximately 1.3%. Figure 4 plots the risk over a range of EOPR values for both micrometers used in the examples.

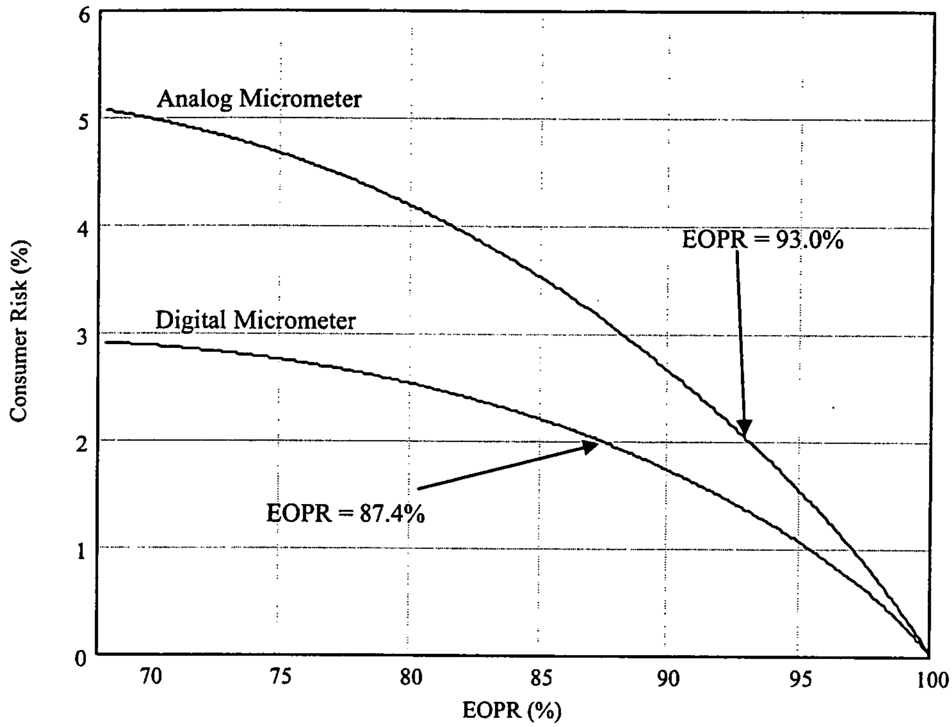

Figure 4: Consumer Risk versus the EOPR.

Summary and Conclusions

It was stated earlier that the devil is in the details. When it comes to the application of requirements, this has never held more truth. As it has been shown, an improperly defined TAR does not have a direct quantitative relationship to measurement decision risk. It was also demonstrated that even an explicitly defined TUR does not provide enough information to assess the suitability of a measurement process. Concise definitions are needed to properly fulfill requirements because "the devil" loves ambiguity.

For over fifty years the TUR/TAR has been a part of metrology and specifically calibration processes. Its development was intended to be a temporary stop-gap due to a lack of mathematical computing power, which is a problem that does not exist today. The goal of this paper was to provide a better understanding of the concepts behind the definitions of TAR and TUR and their relevance to decision risk in measurement processes. With an understanding of these concepts, it is hoped the reader will have a better appreciation of the potential pitfalls of using general rules-of-thumb, especially when the rules are not adequately defined.

Although not discussed in this paper, there are many benefits in using risk analysis tools in lieu of generalized rules. The potential of increasing product or process quality is usually obvious, but there is also the potential for economic benefits. A case in point is the micrometer example; the TURs do not meet the requirements of Z540.1 or Z540.3, but the risk analysis indicates the process has a low Consumer Risk (False Accept Risk). The alternative to risk analysis is to either develop a different calibration process or change the specification tolerances of the micrometer, both which have the potential for negative economic impact. These benefits extend far beyond the discussion of this paper and include other areas such as Producer Risk (False Reject Risk) which can have a large economic impact in the area of rework and product acceptance decisions. The benefits of risk analysis can far out weigh the costs and are limited only by the imagination of the user willing to apply the science.

A final closing thought concerns the value of the risk requirement specified by Z540.3. A requirement for measurement decision risk which "shall not exceed 2%" is probably appropriate for many applications, but not all. In some applications, 2% may be an excessive level of risk (e.g., space flight or nuclear weapons), but what of the cases where 2% risk is not warranted, or a more likely scenario, not obtainable? It is imperative upon users of M&TE to be clear on the requirements for the equipment. The application requirements should establish the *required* measurement decision risk, in lieu of relying on a consensus standard which may not be appropriate, or may not even be obtainable.

It cannot be overstated; the devil is in the details.

References

1. *Metrology – Calibration and Measurement Process Guidelines,* NASA Reference Publication 1342, NASA, 1994
2. Bucher, Jay L (Editor), *The Metrology Handbook,* American Society for Quality, ASQ Quality Press 2004
3. *International Vocabulary of Basic and General Terms in Metrology*, International Organization for Standardization (ISO), 1993
4. *Requirements for the Calibration of Measuring and Test Equipment,* ANSI/NCSL Z540.3-2006, NCSL International, 2006
5. Eagle, Alan R., "A Method for Handling Errors in Testing and Measuring," *Industrial Quality Control*, March 1954.
6. Grubbs, Frank, and Coon, Helen, "On Setting Test Limits Relative to Specification Limits," *Industrial Quality Control*, March 1954
7. Hayes, Jerry L., *Factors Affecting Measuring Reliability,* U.S. Naval Ordnance Laboratory, TM No. 63-106, 24 March 1955
8. Hayes, Jerry L., Series of telephone interviews, January – April, 2007
9. *Safety, Reliability, Maintainability, and Quality Provisions for the Space Shuttle Program,* NSTS 5300.4 (1D-2) Change 9, May 2006, NASA
10. *Handbook for the Interpretation and Application of ANSI/NCSL Z540.1-1994*, NCSL International, 1995
11. *Calibration Laboratories and Measuring and Test Equipment – General Requirements,* ANSI/NCSL Z540.1-1994 (R2002), NCSL International, 2002
12. *American National Standard for Calibration – U.S. Guide to the Expression of Uncertainty in Measurement*, ANSI/NCSL Z540-2-1997 (R2002), NCSL International, 2002
13. Castrup, Howard, *A Note on the Accuracy Ratio Requirements,* Integrated Sciences Group, revised 2006
14. *Uncertainty Analysis Principles and Methods*, Integrated Sciences Group (ISG), 2006

Measurement Decision Risk – The Importance of Definitions

Speaker/Author:
Scott M. Mimbs
National Aeronautics and Space Administration (NASA)
John F. Kennedy Space Center
Kennedy Space Center, FL 32899
Phone: 321-861-5184
Fax: 321-867-1740
scott.m.mimbs@nasa.gov

Abstract

One of the more misunderstood areas of metrology is the Test Uncertainty Ratio (TUR) and the Test Accuracy Ratio (TAR). There have been many definitions over the years, but why are these definitions important to a discussion on measurement decision risk? The importance lies in the clarity of communication. Problems can immediately arise in the application (or misapplication) of the definition of these terms. In other words, while it is important to understand the definitions, it is more important to understand concepts behind the definitions and to be precise in how they are applied.

The objective of any measurement is a decision. Measurement Decision Risk is a way to look at the quality of a measurement, and although it is not a new concept, it has generated a lot of attention since its addition as a requirement in the new U.S. National Standard, ANSI/NCSL Z540.3-2006. In addition to Measurement Decision Risk as the prime method of managing measurement risk, Z540.3 has added, as a fall-back, an explicit definition for TUR. The impact these new requirements may have on calibration service providers has become the topic of much discussion and in some cases concern.

This paper will look at the concepts behind the definitions and how they relate to Measurement Decision Risk. Using common examples, this paper will also provide a comparison of various elements of risk related to measurement science using the concepts of TAR, TUR, accuracy ratios, and Consumer Risk (False Accept Risk). The goal of this paper is to provide a better understanding of their relevance to the measurement decision process.

Introduction

It is well known that the international definition of metrology is the science of measurement [1], but, though metrology may be about measurements, the focus should be on what is being done with the information from the measurements. In the simplest terms, measurements are made to support decisions and/or establish facts. For example, measurements help make decisions:

- To continue or stop a process (including a space launch)
- To accept or reject a product
- To rework or complete a design
- To take corrective action or withhold it
- To establish scientific or legal fact.

- To establish research or investigative fact.

If the data from measurements are not being used in decision-making or in establishing facts (including scientific research), the measurement is unnecessary. *"The more critical the decision, the more critical the data. The more critical the data, the more critical the measurement"* [2].

Validating the suitability of measuring and test equipment (M&TE) is an essential component of the measurement process and calibration is one of the oldest methods of doing this. Calibration has been a part of the measurement process since the Egyptian Cubit Stick was used in the building of the pyramids [3]. As technology increased in the mid-20th century, the need for more precision in testing and calibration led to the development of new methods in determining the risks involved in measurement related decisions. The TUR and TAR developed out of these activities and have been used for over 50 years as tools for mitigating measurement decision risk. However, the problem with the TUR/TAR is the lack of a consensus on how to apply them or even how to define them.

As with any new consensus standard, organizations owning, using, and/or calibrating M&TE will have to review the Z540.3 [4] for potential impacts to their businesses. The review will have to consider impacts that may be either positive or negative. After the review, some organizations will have to make the decision whether or not to incorporate the standard, while other organizations may have to bid on contracts that include the standard. All affected organizations will need to understand the implications of measurement decision risk and the explicit definition of TUR. This paper begins to address the concepts behind measurement decision risk and the TUR by looking at the following areas:

1. A look at the history and development of measurement decision risk and the TUR.
2. Discuss the current definitions of TUR/TAR.
3. Discuss how the Z540.3 definition of TUR is linked to measurement decision risk and the limitations of its use.
4. Review an example of the application of TUR in an off-nominal case and the corresponding calculation of measurement decision risk.

Development of Measurement Decision Risk

A look at the origin of the TUR/TAR helps in understanding their relationship to measurement decision risk.

Measurement decision risk analysis traces its roots to the early work on consumer and producer risk analysis done by Alan Eagle, Frank Grubbs, and Helen Coon [5, 6] in the late 1940's and early 1950's. Eagle's 1954 paper describes the methods for calculating the consumer and producer risk and how to establish "test limits" in relation to design limits which have become known as guardbands today. The focus of the paper was to analyze and mitigate the "test errors" which are "inherent in the test equipment and/or test personnel" used in the inspection of manufactured complex electronic equipment [5]. The key point to Eagle's method was quantifying and using consumer/producer risk (measurement decision risk) as a part of the manufacturing process. This concept is applicable wherever decisions are based on measurements.

In 1955, the U.S. Navy recognized the need for improved measurement reliability in their guided missile program. Building upon Eagle's work [5], Jerry Hayes set out to establish a basis for accuracy ratios versus decision risks for application in the Navy's calibration program [7]. The practice at the time was to use a 10:1 ratio, but that value was considered unsupportable by the nation's calibration support and measurement traceability infrastructure. Using the relationship between the design specifications, testing limits, and instrument error, Hayes proposed using a "family of curves" to determine the specific testing risk or reliability. The problem with this method was a new family of curves had to be established each time a process or design tolerance changed. A change in a process or tolerance nullified the assumptions upon which the first set of curves was built [7, 8]. Some important aspects of the Hayes paper, still relevant today, are the need for calibrated equipment used in testing, establishment of reasonable testing risk levels, reasonable design tolerances, and adequate procedures for testing.

Hayes continued to work on methods of assuring measurement reliability based on consumer risk. In the mid-50's, computing consumer risk was a very arduous task (requiring use of a slide rule), which Hayes decided not to require U.S. Navy contractors to perform. With very specific assumptions on process, a consumer risk of 1% was selected, which calculated to be about a 3:1 accuracy ratio. Hayes, working with Stan Crandon, decided to pad the ratio to account for uncertainty in the reliability of the tolerances industry was using for the measurement standards. Thus the 4:1 ratio requirement was developed and established as Navy policy and subsequently adopted as a requirement in military procurement standards both here and abroad [8]. This ratio became what is known today as the TAR and later evolved into the TUR.

A Look at Existing Definitions

Various definitions for TAR and TUR have appeared in many texts, papers and documents over the years, but there has never been a consensus standard available that provided a definition, much less specified how to apply the rule. The following are some current definitions.

- The American Society for Quality defines TAR and TUR in terms of calibration [3].

  Test accuracy ratio *- (1) In a calibration procedure, the test accuracy ratio (TAR) is the ratio of the accuracy tolerance of the unit under calibration to the accuracy tolerance of the calibration standard used.*

$$\text{TAR} = \frac{\text{UUT\_Tolerance}}{\text{Std\_Tolerance}}$$

  Test uncertainty ratio *- In a calibration procedure, the test uncertainty ratio (TUR) is the ratio of the accuracy tolerance of the unit under calibration to the uncertainty of the calibration standard used.*

$$\text{TUR} = \frac{\text{UUT\_Tolerance}}{\text{Std\_Uncertainty}}$$

  Note: UUT is the Unit under Test and Std represents the calibration standard.

- Although a direct definition of TUR is avoided, NASA's Space Shuttle Program has ratio requirements that apply to calibration and article or material measurements [9].

  Paragraph 4: Article or Material Measurement Processes
  The Expanded Uncertainty in any article or material measurement process shall not exceed ten percent of the tolerance of the article or material characteristic being measured.

Paragraph 5: Calibration Measurement Processes
... the Expanded Uncertainty in any calibration measurement process shall not exceed 25 percent of the tolerance of the parameter being measured.

Note: Expanded Uncertainty is defined in the ANSI/NCSL Z540-2-1997 [10] which is the U.S. adoption of the International Organization for Standardization (ISO) *Guide to the Expression of Uncertainty in Measurement* (GUM).

- ANSI/NCSL Z540.1-1995 [11] is the predecessor to the Z540.3. The Z540.1 Handbook [12], through the use of a note, considered the TAR and TUR interchangeable, although the definition of the TAR in the guidance differs from traditional definitions.

Interpretive Guidance for Section 10.2 of the Handbook [12]

As a default alternative to doing an uncertainty analysis, a laboratory may rely on a Test Accuracy Ratio (TAR) of 4:1. A TAR of 4:1 means that the tolerance of the parameter (specification) being tested is equal to or greater than four times the combination of the uncertainties of all the measurement standards employed in the test.

If it is determined that the TAR is less than 4:1, then one of the following methods may be used: uncertainty analysis as described above, guard-banding, widening the specification, or another appropriate method.

Note: Some refer to TARs as Test Uncertainty Ratios or TURs

- The Z540.3 [4] provides an explicit definition of TUR, but does not address the TAR.

3.11 Test uncertainty ratio
The ratio of the span of the tolerance of a measurement quantity subject to calibration, to twice the 95% expanded uncertainty of the measurement process used for calibration.

NOTE: This applies to two-sided tolerances.

The definition uses the expanded uncertainty as defined in Z540-2 [10] where *k* is the coverage factor. The definition in equation form:

$$\mathrm{TUR} = \frac{\mathrm{Upper} - \mathrm{Lower}}{2 \cdot \mathrm{U}} \qquad \mathrm{U} = \mathrm{k} \cdot \mathrm{u} \qquad \mathrm{k} = 2$$

There are key differences between the Z540.3 and earlier TUR definitions.

1. The earlier TUR denominator is not well defined which leads to inconsistent applications.
2. The denominator for the Z540.3 TUR is explicitly defined, thus providing better uniformity in the application of the TUR.

The ASQ definition of the TAR [3] is one of the more popular applications of the risk rules-of-thumb, most likely due to the simplicity of its implementation. It is used in many labs and other applications and can be found in many papers and training guides for calibration and quality inspection. It must be cautioned however, as such, the ASQ definition of the TAR is non-compliant with either Z540.1 or Z540.3 and can lead the user into a false sense of security, which is a main focus of this paper

As mentioned earlier, the objective of a measurement is a decision and Measurement Decision Risk is a tool to assess the suitability of the measurement process. The intent of a TAR or a TUR is to mitigate the risks in measurement, but their value is diminished without a concise definition. Problems can quickly arise when different meanings are applied to the same term. The devil is in the details.

Linking the TUR to Measurement Decision Risk

It is from Eagle's work the TUR was first developed [7, 8]. The consumer risk equation associated with Eagle's work [5, 6] is presented here, but will not be discussed in detail.

$$CR = \frac{1}{\pi}\cdot\int_{k}^{\infty}\int_{-r\cdot(k+t)+b}^{r\cdot(k-t)-b} e^{\frac{-\left(t^2+s^2\right)}{2}}\,ds\,dt$$

To help explain the relationship of the TUR to Measurement Decision Risk, the functional relationship of the equation will be discussed using the three "influence" variables, *b*, *r*, and *k*.

$$CR = CR(b, k, r)$$

The variable *b* is the deviation of the test limits from the specification limits [i.e., $\mu \pm (k\sigma_X - b\sigma_e)$]. This establishes the required "test limit,"(also known as a guardband) to achieve a desired consumer risk. For the purposes of this discussion, the specification limits will equal the test limits (i.e., $b = 0$), therefore leaving only two influence variables for this discussion.

The variable *k* is the number of standard deviations the performance specification limit is to the product mean, which is assumed to be centered (i.e., $\mu \pm k\sigma_X$). Figure 1 illustrates the relationship of the specification limits and production (or process) distribution with the limits falling at $\pm 2\sigma$.

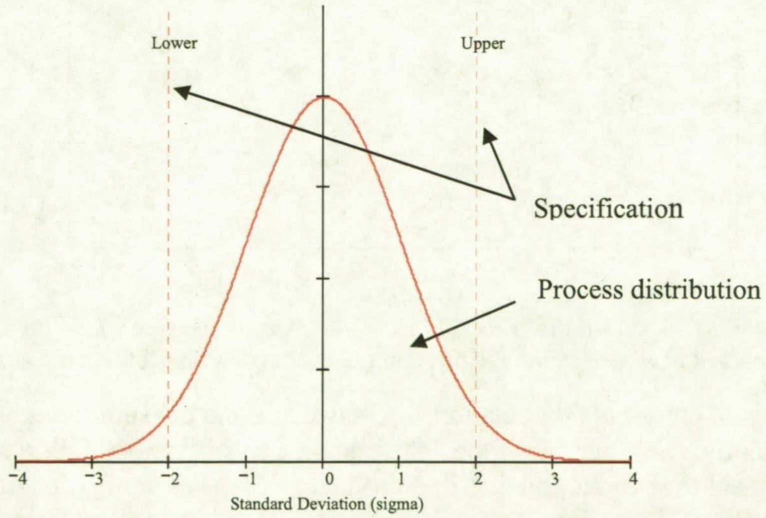

Figure 1: The relationship of the specification limits to the process distribution for ±2σ.

The last variable to consider is r which is a ratio of two standard deviations.

$$r = \frac{\sigma_X}{\sigma_e}$$

The numerator σ_X is what Eagle described as "the true standard deviation of the product distribution." This represents the subject of interest. Eagle described the denominator σ_e as "the standard deviation of the errors of measurement" [5]. In metrology, the standard deviation is an indicator of uncertainty (i.e., $\pm k\sigma$ indicates $\pm k$u). Although it is considered the origin of the TUR and is referred to as an "accuracy ratio," without any confidence limits or coverage factors ($\pm k$), in reality, r is a pure "uncertainty ratio."

$$r = \frac{\sigma_X}{\sigma_e} = \frac{u_X}{u_e}$$

The relationship between r and k becomes more evident when graphically displayed. Figure 2 shows the "knee" of the curve intersecting the 1% risk level at approximately an r-value of 3:1. As 1% was the risk level Hayes was seeking, adding a little extra margin, 4:1 became the "rule of thumb" [7, 8].

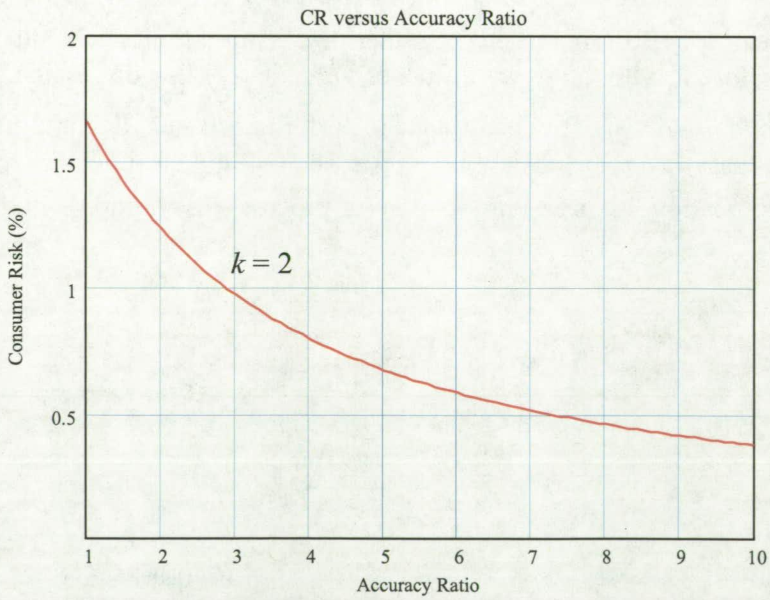

Figure 2: Consumer Risk versus the "accuracy ratio" for the specific case where the specification limits lie at 2 standard deviations from the mean of the production distribution (i.e., k=2 for $\mu \pm 2\sigma_X$).

The curve in Figure 2 represents the relationship between r and consumer risk where k = 2. Eagle noted in his original paper that if the "performance specification" falls at more than 2 standard deviations of the production distribution (i.e., $k > 2$), then mitigation efforts are not required [5]. His statement may not be true for all cases in today's technology, but as discussed earlier, this was the assumption that Hayes and Crandon used to establish the 4:1 rule in the mid-1950s [8].

Hayes and Crandon used a chart similar to Figure 2 for their analysis as they searched for a method to improve measurements without levying a large amount of calculations on the contractors. When Hayes allowed the use of a ratio between the tolerances of the subject of interest and the measuring equipment, the idea was supposed to be temporary until better computing power became available or a better method could be developed [8]. As discussed earlier, Hayes settled on 4:1, but others went with a more conventional and conservative 10:1. NASA used the 10:1 for all calibration and article measurement requirements through the first moon landing in 1969. After that, calibration requirements were changed to 4:1 while test measurement requirements remained at 10:1 [9].

As time has gone by, many have tried to help the TUR/TAR become a tool that is more soundly founded in measurement science. The Z540.3 comes very close to accomplishing that goal, but as will be demonstrated, more information is required to use the TUR. To explain this, let's take another look at the ratio from the consumer risk equation, but now only extend the standard deviation to measurement process uncertainty.

$$r = \frac{\sigma_X}{\sigma_e} = \frac{\sigma_X}{u_e}$$

Remember that the relationship of the tolerance to the product mean is $\mu \pm k\sigma_X$. Combined with the value of r as stated above, the Z540.3 TUR can be transformed (slightly) where the relationship of $\pm k\sigma_X$ becomes more evident.

$$\text{TUR} = \frac{\text{Upper} - \text{Lower}}{2 \cdot U_{95}} = \frac{\text{Upper} - \text{Lower}}{2 \cdot k_e \cdot u_e} = \frac{2 \cdot k_X \cdot \sigma_X}{2 \cdot k_e \cdot u_e} = \frac{k_X \cdot \sigma_X}{k_e \cdot u_e}$$

As shown here, the TUR can be represented as a ratio of intervals ($\pm k\sigma$). For Z540.3, the coverage factor in the denominator is usually inferred to be $k = 2$ (95% of normal). Although there is no definition for the value of k in the numerator, it might be assumed that it represents the k factor in the consumer risk equation. It is this k-factor that becomes the link between the TUR and measurement decision risk.

The next step is to establish a method for obtaining the value of the k-factor for a calibration process. One method that can be used for calibration is to allow the End of Period Reliability (EOPR) for the Unit under Test (UUT) to represent the "product distribution" [13]. EOPR is the probability of a unit being in-tolerance when it is returned for routine calibration at the end of its normal interval. The EOPR is reported in percentages and if the EOPR is assumed to be normally distributed, this percentage can be related back to a confidence limit (or coverage factor). For example, an EOPR of 95.45% would then be represented by a value of $k = 2.0$. It is in this manner that the EOPR fits the relationship of specification limits to the process distribution as previously shown in Figure 1. For this type of relationship, the value of k can be obtained from statistic books or calculated.

Using EOPR to represent the UUT product distribution opens up many questions about the level of "drill down" required in obtaining this information (e.g., nomenclature, manufacture, model number, serial number, or parameter value). The answers to these questions are beyond the scope of this paper, but an excellent detailed discussion can be found in Section 6.3 of NASA Reference Publication 1342 [2].

Applying Definitions – Example 1

A generic example of a UUT and a working standard will be used to illustrate the concepts just discussed. The uncertainty is a generic estimate with no assumptions defined, as it only represents the combined standard uncertainty of the measurement process

Parameter	UUT	Working Standard
Tolerance	±4.0	±1.0
$u_c = 0.67$		
TAR = 4:1		
TUR = 3:1		

Table 1: A generic calibration example.

With a TUR of less than 4:1, if the Z540.3 is applicable, this calibration would be unacceptable for application of the TUR fall-back rule. But is the calibration "good enough" – is the Measurement Decision Risk acceptable (2% or less per the Z540.3)?

As discussed above, the assumption in the creation of the original 4:1 ratio, with respect to the desired level of measurement risk, was 1% [8]. The value of *k* was a pivotal assumption because when $k = 2$ (in a normal distribution), more than 95% of the process would be contained within the specification limits. As stated earlier, the product distribution is to be represented by the EOPR of the UUT. Figure 3 illustrates the relationship between consumer risk and EOPR for TUR ratios of 4:1 and 3:1.

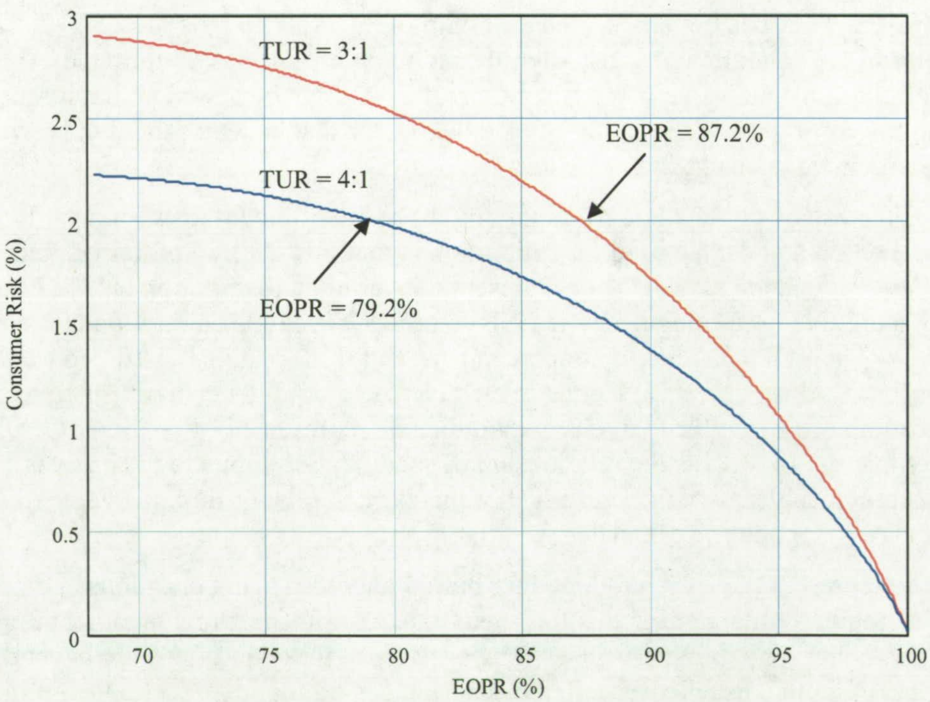

Figure 3: Consumer Risk versus the EOPR.

Figure 3 shows it is possible for a TUR below 4:1 to be acceptable (given 2% is acceptable), but it also demonstrates that 4:1 may not provide a desired level of risk. For example, as Figure 3 shows, the EOPR must be above 79% for a 4:1 to provide a consumer risk of 2%. The bottom-line is a TUR may not provide a proper or desired level of quality without additional information. As stated earlier, using the TUR alone may lull an organization into a false sense of security by not quantifying a key component of risk.

Example 2

All types of calibrations will be affected by Z540.3, but those that do not fit the general mold of the TUR or 2% rule are the calibrations that will create the most problems for quality mangers and calibration service providers. This example will look at the calibration of a digital micrometer, an operation performed millions of times a year across our planet. Table 2 contains the specification information for the micrometer and calibration standard (Class 2 Gage Blocks).

Parameter	Digital Micrometer 0-1 inch (0-25.4 mm)	Gage Block, Class 2
Tolerance	± 0.0001 inch (± 2.54 μm)	± 0.000004 inches (± 0.1016 μm)
Resolution	± 0.00005 inch (± 1.27 μm)	N/A

Table 2: The specification information for the digital micrometer and the calibration standard.

A new or overhauled micrometer is calibrated using at least 10 points to cover its range, but only 4 points for routine calibrations (0, plus low, middle, and high). The purpose of this example is to examine the measurement risk using TAR and TUR, therefore only one point of the calibration will be discussed - the mid point of 0.50000 inches (12.700 mm).

Using the ASQ definition, the calibration TAR has a value of 25:1.

$$\text{TAR} = \frac{\text{UUT\_tolerance}}{\text{Std\_tolerance}} = \frac{0.0001}{0.000004} = 25$$

As previously stated, the purpose of this example (and paper) is the examination of usage of measurement risk tools. As such, the uncertainty analysis is presented in table form without the detailed calculations. The analysis follows the guidelines of the Z540-2 [10]. The specific methods followed were those recommended in *Uncertainty Analysis Principles and Methods* [14], developed for the Naval Air Warfare Center Aircraft Division (NAWCAD) and Air Force Flight Test Center (AFFTC). Readers desiring the detailed calculations may contact the author. Table 3 contains the analysis results for the digital micrometer.

Uncertainty Source	Standard Uncertainty inches (µm)	Confidence Level (%)	Type (A or B)	Distribution
Gage Blocks	0.000002 (0.0508)	95.00	B	Normal
Resolution	0.0000144 (0.3658)	100.00	B	Uniform
Environmental	0.0000001 (0.00254)	95.00	B	Normal
Random Error or Repeatability	0.0000083 (0.2108)	95.00	A	Student's t
Combined Uncertainty	0.0000168 (0.4267)			

Table 3: Results of the uncertainty analysis for a digital micrometer.

The combined standard uncertainty from Table 3 is calculated into the Z540.3's TUR for the following result. For the 95% expanded uncertainty, assume $k = 2$.

$$\mathrm{TUR} = \frac{\mathrm{Upper} - \mathrm{Lower}}{2 \cdot U_{95}} = \frac{0.0002}{2 \cdot 2 \cdot 0.0000168} = 3$$

As can be seen, even with a very high TAR, when the calibration standard is passive (i.e., the UUT is the "reading" instrument), the TUR may fall below 4:1, in this case due to instrument resolution.

Example 3

The subject of Example 2 was a digital micrometer with a resolution that is one half of the unit tolerance. Many times the micrometer resolution and tolerance are equal, as in the case for Example 3. In this example, the micrometer has a vernier scale in addition to the regular graduations, allowing the user to read more precisely. The TAR for this example is 25:1, the same as for Example 2. Table 4 contains the specification information for this example.

Parameter	Analog (Vernier) Micrometer 0-1 inch (0-25.4 mm)	Gage Block, Class 2
Tolerance	± 0.0001 inch (± 2.54 µm)	± 0.000004 inches (± 0.1016 µm)
Resolution	± 0.0001 inch (± 2.54 µm)	N/A

Table 4: The specification information for the analog micrometer and the calibration standard.

Uncertainty Source	Standard Uncertainty inches (µm)	Confidence Level (%)	Type (A or B)	Distribution
Gage Blocks	0.000002 (0.0508)	95.00	B	Normal
Resolution	0.0000255 (0.6477)	95.00	B	Normal
Environmental	0.0000001 (0.00254)	95.00	B	Normal
Random Error or Repeatability	0.0000201 (0.5105)	95.00	A	Student's t
Combined Uncertainty	0.0000326 (0.8280)			

Table 5: Results of the uncertainty analysis for the analog micrometer.

The results of the uncertainty analysis are taken from Table 5 and again calculated into the Z540.3's TUR for the following result.

$$\mathrm{TUR} = \frac{\mathrm{Upper} - \mathrm{Lower}}{2 \cdot U_{95}} = \frac{0.0002}{2 \cdot 2 \cdot 0.0000326} = 1.5$$

As expected, the results are well below the 4:1 TUR required by both the Z540.1 and Z540.3.

Discussion of Example Results

The examples above were chosen for a reason – they do not meet the requirements of the existing or new standards for calibration in terms of the TUR. But do they meet the Z540.3's 2% rule for Measurement Decision Risk? The best way to answer that question is to calculate the consumer risk for the calibration. For illustrative purposes, the consumer risk for only the analog micrometer will be calculated.

The consumer risk equation will be set as a function of the variables *k*, *r*, and *b*.

$$\mathrm{CR}(k,r,b) = \frac{1}{\pi} \cdot \int_{k}^{\infty} \int_{-r\cdot(k+t)+b}^{r\cdot(k-t)-b} e^{\frac{-\left(t^2+s^2\right)}{2}} \, ds \, dt \qquad r = \frac{\sigma_x}{u_e}$$

At one of Kennedy Space Center's calibration labs, the nomenclature "micrometer" has an EOPR of just over 96%, but the model number of the analog micrometer has an EOPR of 97.3%. For an assumed normal distribution, this equals a coverage factor of 2.21 (i.e., $k = 2.21$).

Using the EOPR data to solve for the value of σ_x and *r*, then inserting the information into the function provides the following results.

$$\sigma_X = \frac{0.0001}{2.21} = 0.0000452 \quad r = \frac{0.0000452}{0.0000326} = 1.387$$

Consumer Risk: $CR(2.21, 1.387, 0) = 0.9\%$

If the model specific EOPR information were not available, then the overall nomenclature EOPR of 96% would have yielded a consumer risk for the calibration of approximately 1.3%. Figure 4 plots the risk over a range of EOPR values for both micrometers used in the examples.

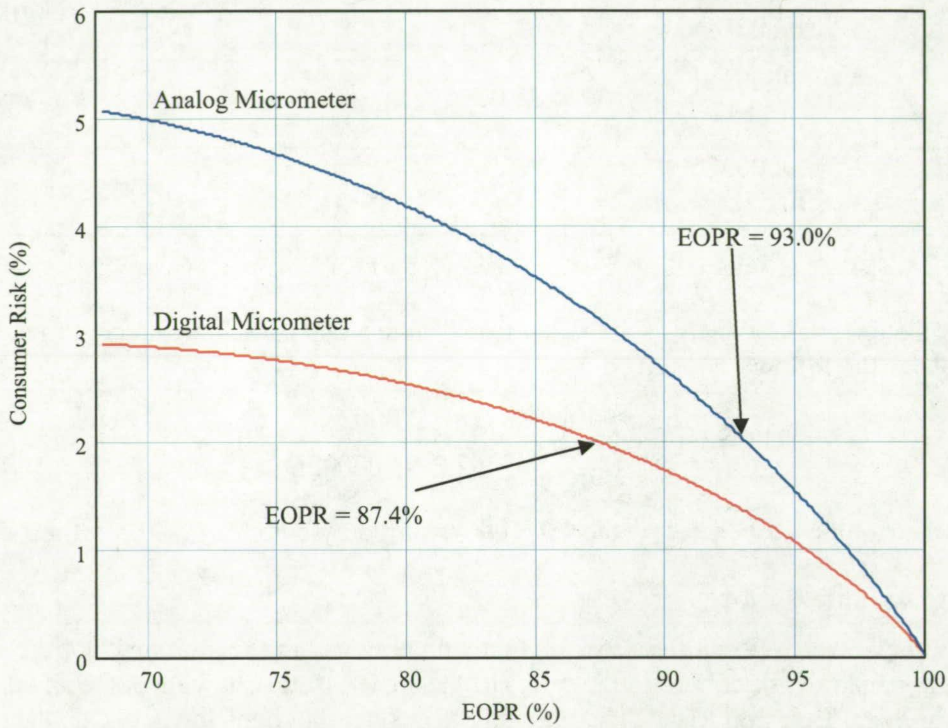

Figure 4: Consumer Risk versus the EOPR.

Summary and Conclusions

It was stated earlier that the devil is in the details. When it comes to the application of requirements, this has never held more truth. As it has been shown, an improperly defined TAR does not have a direct quantitative relationship to measurement decision risk. It was also demonstrated that even an explicitly defined TUR does not provide enough information to assess the suitability of a measurement process. Concise definitions are needed to properly fulfill requirements because "the devil" loves ambiguity.

For over fifty years the TUR/TAR has been a part of metrology and specifically calibration processes. Its development was intended to be a temporary stop-gap due to a lack of mathematical computing power, which is a problem that does not exist today. The goal of this paper was to provide a better understanding of the concepts behind the definitions of TAR and TUR and their relevance to decision risk in measurement processes. With an understanding of these concepts, it is hoped the reader will have a better appreciation of the potential pitfalls of using general rules-of-thumb, especially when the rules are not adequately defined.

Although not discussed in this paper, there are many benefits in using risk analysis tools in lieu of generalized rules. The potential of increasing product or process quality is usually obvious, but there is also the potential for economic benefits. A case in point is the micrometer example; the TURs do not meet the requirements of Z540.1 or Z540.3, but the risk analysis indicates the process has a low Consumer Risk (False Accept Risk). The alternative to risk analysis is to either develop a different calibration process or change the specification tolerances of the micrometer, both which have the potential for negative economic impact. These benefits extend far beyond the discussion of this paper and include other areas such as Producer Risk (False Reject Risk) which can have a large economic impact in the area of rework and product acceptance decisions. The benefits of risk analysis can far out weigh the costs and are limited only by the imagination of the user willing to apply the science.

A final closing thought concerns the value of the risk requirement specified by Z540.3. A requirement for measurement decision risk which "shall not exceed 2%" is probably appropriate for many applications, but not all. In some applications, 2% may be an excessive level of risk (e.g., space flight or nuclear weapons), but what of the cases where 2% risk is not warranted, or a more likely scenario, not obtainable? It is imperative upon users of M&TE to be clear on the requirements for the equipment. The application requirements should establish the *required* measurement decision risk, in lieu of relying on a consensus standard which may not be appropriate, or may not even be obtainable.

It cannot be overstated; the devil is in the details.

References

1. *International Vocabulary of Basic and General Terms in Metrology*, International Organization for Standardization (ISO), 1993
2. *Metrology – Calibration and Measurement Process Guidelines*, NASA Reference Publication 1342, NASA, 1994
3. Bucher, Jay L (Editor), *The Metrology Handbook,* American Society for Quality, ASQ Quality Press 2004
4. *Requirements for the Calibration of Measuring and Test Equipment,* ANSI/NCSL Z540.3-2006, NCSL International, 2006
5. Eagle, Alan R., "A Method for Handling Errors in Testing and Measuring," *Industrial Quality Control*, March 1954
6. Grubbs, Frank, and Coon, Helen, "On Setting Test Limits Relative to Specification Limits," *Industrial Quality Control*, March 1954
7. Hayes, Jerry L., *Factors Affecting Measuring Reliability,* U.S. Naval Ordnance Laboratory, TM No. 63-106, 24 March 1955
8. Hayes, Jerry L., Series of telephone interviews, January – April, 2007
9. *Safety, Reliability, Maintainability, and Quality Provisions for the Space Shuttle Program,* NSTS 5300.4 (1D-2) Change 9, May 2006, NASA
10. *American National Standard for Calibration – U.S. Guide to the Expression of Uncertainty in Measurement*, ANSI/NCSL Z540-2-1997 (R2002), NCSL International, 2002
11. *Calibration Laboratories and Measuring and Test Equipment – General Requirements,* ANSI/NCSL Z540.1-1994 (R2002), NCSL International, 2002
12. *Handbook for the Interpretation and Application of ANSI/NCSL Z540.1-1994*, NCSL International, 1995
13. Castrup, Howard, *A Note on the Accuracy Ratio Requirements,* Integrated Sciences Group, revised 2006
14. *Uncertainty Analysis Principles and Methods*, Integrated Sciences Group (ISG), 2006