The new dimension to resolution: Can it be resolved?

Author: Henry A Zumbrun, Company: Morehouse Instrument Company, hzumbrun@mhforce.com

Co-Author: Dilip Shah, E=mc3 solutions, emc3solu@aol.com

Abstract

In the metrology community, there is an ongoing debate over which contributors to the Unit Under Test (UUT) belong in the expanded uncertainty calculation of the measurement process used for calibration. This is also known as Calibration Process Uncertainty (CPU); CPU is the denominator when calculating a Test Uncertainty Ratio (TUR).

$$\mathbf{TUR} = \frac{\textbf{Span of the } \pm \textbf{UUT Tolerance}}{\mathbf{2\ x\ k_{95\%}}(\textbf{Calibration Process Uncertainty})}$$

Figure 1: TUR Formula found in ANSI/NCSLI Z540.3 Handbook

This paper presents examples that illustrate why the best practices outlined in documents such as ILAC-P14:09/2020 and the ANSI/NCSLI Z540.3 Handbook should be followed regarding the contributors for the CPU. Instead of drafting their own test protocols and standards, calibration laboratories and manufacturers are advised to correctly calculate both uncertainty and risk. Performing these calculations is part of an ethical approach to calibration that avoids shifting more risk to the Industry and ultimately mitigates global consumer's risk. Furthermore, outdated approaches to calculations, such as Test Accuracy Ratio (TAR), must be discontinued, and efforts to change the agreed-upon definition of Test Uncertainty Ratio (TUR) should cease since modern computing can provide measurements that are more accurate and reliable.

Definition and Calculation of Test Uncertainty Ratio (TUR)

Understanding Test Uncertainty Ratio (TUR) is the first step in weighing the significance of the claims proposed in this research paper. TUR is defined as:

- 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.[1]
- The ratio of the tolerance, TL, of a measurement quantity, divided by the 95% expanded measurement uncertainty of the measurement process where TUR = TL/U. [2]

These definitions are similar, but the span of the tolerance in the numerator must be clearer. If the tolerance is not symmetrical, then ANSI/NCSL Z540.3 is much clearer.

The calculation of TUR is crucial because it is a commonly accepted practice in making a statement of conformity. When used in combination with the measurement location, one can calculate measurement risk at the time of calibration. TUR is clearly defined in standards such as ANSI/NCSL Z540.3 and the ANSI/NCSL Z540.3 Handbook. However, equipment manufacturers participate in the practice of writing unique standards that favorably market their products. In reality, the application of the products may be varied, and acceptance requirements may change depending on application. Therefore, the failure to adhere to clearly defined, universal standards can put customers or consumers at increased risk.

Accreditation bodies' recommended requirements on the contributors to Measurement Uncertainty should be considered by the end-user to aid in calculating the optimal TUR. One may argue that the reference standard uncertainty and environmental factors are the only requirements needed in the TUR calculation to make a conformity assessment decision. Make no mistake: TUR is defined and agreed upon in ANSI/NCSL Z540.3:2006 and the ANSI/NCSL Z540.3 Handbook. Therefore, it should not be a point of debate.

The definition of CPU establishes whether relevant uncertainty contributors of the customer's device will be considered by the calibration laboratory that is calibrating the equipment. If a calibration laboratory does not include these uncertainty contributors, then they are passing the risk on to the customer or consumer because the laboratory prefers not to retain the risk. This is often done without the end-user's knowledge.

Evaluating Global Consumer Risk

The customer or consumer is likely a company making a measurement that could have an impact on public safety. Henry Petroski addresses this issue in his book *To Engineer is Human: The Role of Failure in Successful Design*: "Failures appear to be inevitable in the wake of prolonged success, which encourages lower margins of safety. Engineers and the companies who employ them tend to get complacent when things are good; they worry less and may not take the right preventative actions."[3] Petroski's statement about complacency may describe what is happening in the metrology community today regarding the evaluation of global consumer's risk.

Global consumer's risk is defined in JCGM 106:2012. The role of CPU in conformity assessment is defined as "the probability that a non-conforming item will be accepted based on a future measurement result."[4]

When laboratories loosen the restraints or fail to capture the proper contributors while calculating CPU to make a statement of conformity, they are creating a risk wherever this equipment will be used. The application could be weighing an aircraft or overhead material handling, where lives are at stake.

How Calibration Chain Hierarchy Impacts Risk Propagation

If only one tier of the calibration chain cares about the measurement decision risk, then the whole process is at risk. When this risk is propagated throughout succeeding tiers, can we expect the process to work properly?

Metrological traceability is defined as the "property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty."[5] The International System of Units (SI) is at the top of the measurement hierarchy pyramid.

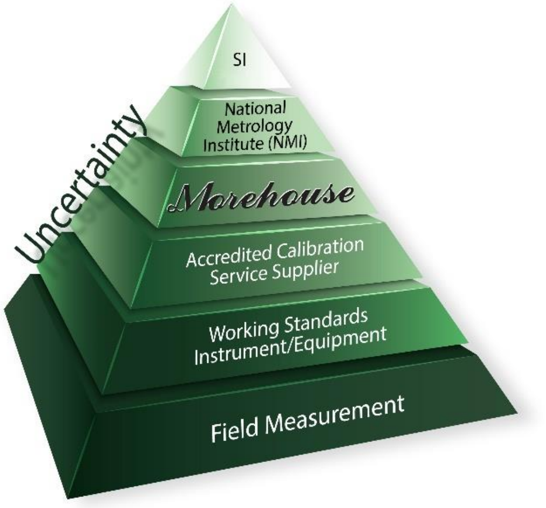


Figure 2: Metrological Traceability Pyramid for Force Measurement

The next tier in the pyramid is a National Metrology Institute (NMI), such as the National Institute of Standards and Technology (NIST) and other designated institutes recognized under the CIPM MRA. In the example above, the third tier is a commercial laboratory, such as Morehouse Instrument Company, with primary standards for force and torque. The lower tiers are accredited calibration suppliers, followed by working standards, with field measurement at the bottom.

In an ideal world, each of these tiers should use the same methods to calculate CPU. Suppose any tier in this pyramid uses a different formula for CPU or neglects critical contributors to the measurement uncertainty. In that case, the next tier will under-report the measurement uncertainty, thereby increasing the overall risk of product failures (PFA – Probability of False Accept).

Considerations for Evaluating Measurement Uncertainty

To meet the metrological traceability requirements, measurement uncertainty must be properly evaluated, taking into consideration the minimum number of contributors. Measurement uncertainty is defined as a "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used."[6]

In simplistic terms, measurement uncertainty may be thought of as doubt or, in effect, doubting the validity of the measurement. CPU is the non-bias uncertainty ascribed to the result of a measurement at a particular test point. It is crucial for calculating the Test Uncertainty Ratio (TUR), as shown in Figure 1.

Most likely, a general industry manufacturer will not calculate the measurement uncertainty associated with their measuring equipment. Although they bear some responsibility, they may not understand their equipment or how to calculate measurement uncertainty. The metrology community has spent decades putting proper measurement uncertainty practices in place. However, the general industry manufacturer is lagging decades behind these practices.

For example, many general industry manufacturers are still using outdated concepts, such as Test Accuracy Ratio (TAR) and NIST traceability. The metrology community has moved away from these terms, but both are still implemented instead of sound metrological practice.

Metrological traceability relies on tracing the measurement chain back to SI units. The proper way to claim metrological traceability is to trace each calibration back to the SI through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. The chain will often end with an NMI such as NIST, which is recognized under the CIPM MRA. Assuming that the end-user will understand the risk seems like a rationalization for equipment manufacturers to sell more products and increase profit margins while knowing full well that the burden of risk is thrust upon the end-user.

Many do not agree on which contributors should be included in a measurement uncertainty evaluation. Some argue that the UUT's resolution does not need to be included in an uncertainty evaluation. However, if the uncertainty evaluation does not consider known contributors to measurement uncertainty, there is a question of whether the calibration can even be metrologically traceable or accredited.

In such a scenario, a calibration laboratory cannot make a statement of conformity or "Pass" an instrument without violating ISO/IEC 17025:2017, which states, "When a statement of conformity to a specification or standard is provided, the laboratory shall document the decision rule employed, taking into account the level of risk (such as false accept and false reject and statistical assumptions) associated with the decision rule employed, and apply the decision rule."[7]

The United Kingdom Accreditation Service (UKAS) elaborates on this point: "Conformity statements under ISO/IEC 17025:2017 require a Decision Rule (3.7) that takes account of measurement uncertainty. Some people argue that it is possible to 'take account' by ignoring it if that is what the customer requests; however, this seems to require a rather contradictory belief that you can be 'doing something' by 'not doing something' (is it possible to 'obey a red stoplight' by 'not obeying a red stoplight'?)."[8]

The UUT resolution is a relevant short-term contributor for consideration in the measurement process uncertainty to the customer's device, and it should not be ignored.

Example #1: Comparing Resolution for a Dimensional Measurement

Think about a user making a measurement with a caliper and a micrometer with different resolutions. In Figure 3, someone is using a caliper to make a measurement. The drawing specification is ±0.000 5 in., and the user has a caliper and wants to know if the part is within tolerance.

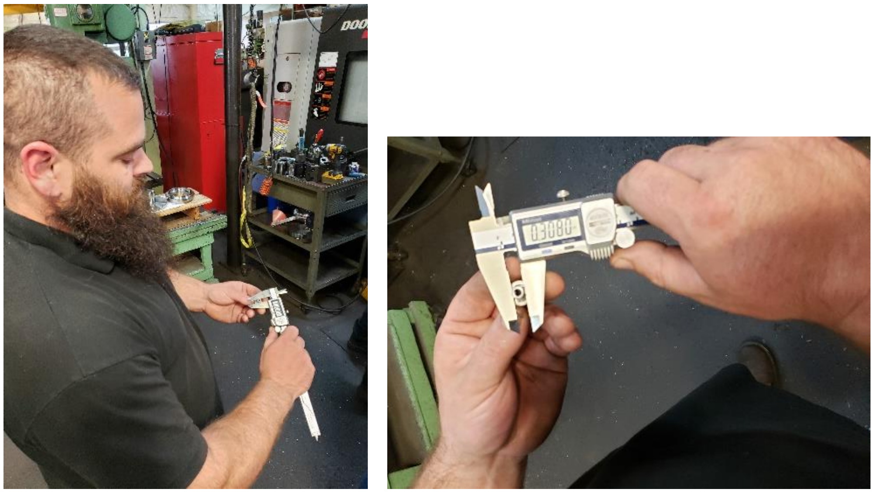


Figure 3: User measuring a part with a caliper with a resolution of 0.000 5 in.

Suppose the laboratory that calibrated the caliper ignored measurement uncertainty contributors, such as the device's resolution and repeatability. Would the user know to shrink the acceptance limit on the drawing so they can say with little doubt that the part is in tolerance? Is the device with a resolution of 0.000 5 in. good enough for the operator to call the part in tolerance? Would the experienced operator instead choose a micrometer to measure the part? The answer to all three of these questions is the same: not necessarily.

In Figure 4, the same item is measured with a micrometer that is capable of reading 0.000 1 in. In one instance, the reading is 0.308 0 in., and in another, the reading 0.308 3 in.

Figure 4: User measuring a part with a micrometer with a resolution of 0.000 1 in.

Example #2: Comparing Resolution for a Weight Measurement

Consider a crane scale that is used to measure the weight of an object. Suppose the laboratory that calibrated the crane scale based the decision rule only on the calibration equipment and not on the crane scale's repeatability or resolution. Would the crane scale operator know to guard band the tolerance at the point of use?

The user has a scale that is known to read with a resolution of 0.1 kg. They need to measure 1 000 kg of uranium to within ± 0.1 kg. Will they guard band to keep what they are weighing within 0.07 kg? Figure 5 illustrates how the user would guard band by subtracting the measurement uncertainty.

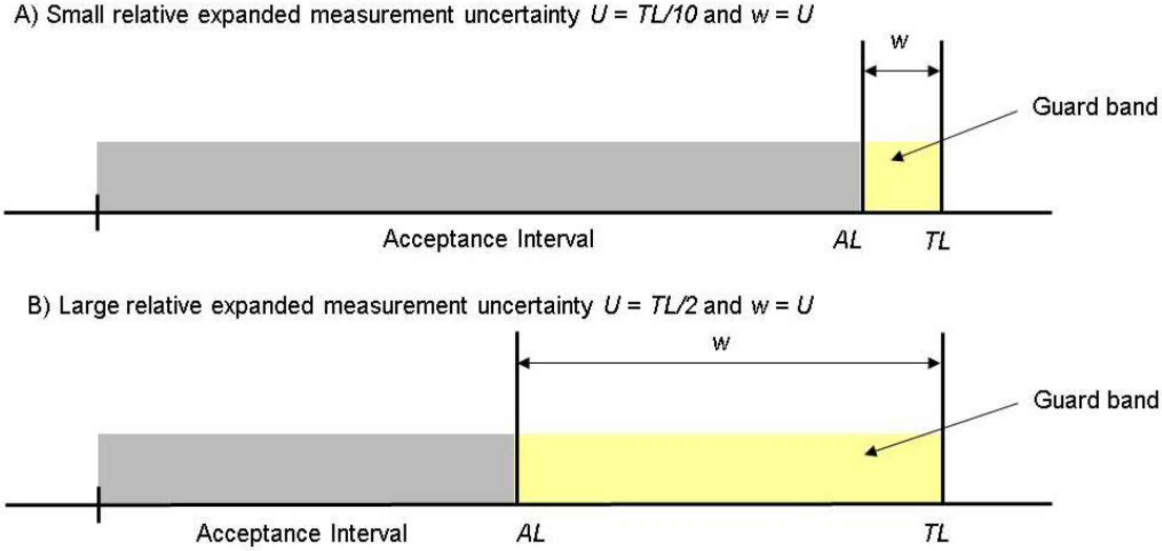


Figure 5: Guard band and acceptance interval illustration found in ILAC G8:09/2019

Risk, Conformity Assessment, and the Decision Rule

These examples emphasize how a device's resolution impacts risk. ISO/IEC 17025:2017 addresses how to account for risk: "When a statement of conformity to a specification or standard for test or calibration is provided, the laboratory shall document the decision rule employed, taking into account the level of risk (such as false accept and false reject and statistical assumptions) associated with the decision rule employed and apply the decision rule."[9]

Some end-users may not understand the decision rule and what it means for them. Are they comfortable with a conformity assessment that does not include all necessary uncertainty contributors, such as the device's resolution? Do they only want a statement that says "Pass" with a sticker that allows them to use the device to make measurements? Are they aware of the potential risks?

Perhaps the end-user is well versed in risk and chooses to follow ANSI/NCSL Z540.3 clause 5.3b. Maybe their purchase order requests that the ANSI/NCSL Z540.3 standard be followed and the decision rule specified. This would be the better—or safer—scenario, but it is not the guaranteed scenario.

Contributors in the TUR Calculation

Many decision rules require a TUR calculation. The formula's ratio includes a numerator and a denominator. ANSI/NCSL describes, "For the numerator, the tolerance used for Unit Under Test (UUT) in the calibration procedure should be used in the calculation of the TUR. This tolerance is to reflect the organization's performance requirements for the Measurement & Test Equipment (M&TE), which are, in turn, derived from the intended application of the M&TE. In many cases, these performance requirements may be those described by the Manufacturer's tolerances and specifications for the M&TE and are therefore included in the numerator."[10]

$$\mathbf{TUR} = \frac{\textbf{Span of the } \pm \textbf{ Tolerance}}{\mathbf{2\ x\ k_{95\%}}\left(\sqrt[2]{\left(\frac{\mathbf{CMC}}{\mathbf{k_{CMC}}}\right)^{2} + \left(\frac{\mathbf{Resolution_{UUT}}}{\sqrt[2]{\mathbf{12}}}\right)^{2} + \left(\frac{\mathbf{Repeatability_{UUT}}}{\mathbf{1}}\right)^{2} + \cdots (\mathbf{u}_{Other})^{2}}\right)}$$

Figure 6: Example of a TUR Formula (Adapted from the ANSI/NCSL Z540.3 Handbook)

In most cases, the numerator is the UUT Accuracy Tolerance. The denominator is slightly more complicated. Per the ANSI/NCSL Z540.3 Handbook, "For the denominator, the 95 % expanded uncertainty of the measurement process used for calibration following the calibration procedure is to be used to calculate TUR. The value of this uncertainty estimate should reflect the results that are reasonably expected from the use of the approved procedure to calibrate the M&TE. Therefore, the estimate includes all components of error that influence the calibration measurement results, which would also include the influences of the item being calibrated except for the bias of the M&TE. The calibration process error, therefore, includes temporary and non-correctable influences incurred during the calibration such as repeatability, resolution, error in the measurement source, operator error, error in correction factors, environmental influences, etc."[11]

This definition of the TUR denominator aligns very closely with ILAC P14:09/2020, which states, "Contributions to the uncertainty stated on the calibration certificate shall include relevant short-term contributions during calibration and contributions that can reasonably be attributed to the customer's device. Where applicable, the uncertainty shall cover the same contributions to uncertainty that were included in evaluation of the CMC uncertainty component, except that uncertainty components evaluated for the best existing device shall be replaced with those of the customer's device. Therefore, reported uncertainties tend to be larger than the uncertainty covered by the CMC."[12]

The TUR formula in Figure 6 is an adaptation with the denominator clarified for current practices. Some may contend that resolution is accounted for with repeatability studies. However, if repeatability is equal to zero, then the UUT's resolution must be considered.

ILAC P14: 09/2020 addresses when the UUT's resolution needs to be included by stating, "When it is possible that the best existing device can have a contribution to uncertainty from repeatability equal to zero, this value may be used in the Evaluation of the CMC. However, other fixed uncertainties associated with the best existing device shall be included."[13]

To correctly calculate TUR, many in the metrology community believe the formula in Figure 6 comprises the minimum contributors that should be included in the denominator for CPU. The formula includes the ratio of UUT Accuracy Tolerance, which manufacturers often request as the accuracy specification, compared against the expanded uncertainty of the calibration process.

At a minimum, the expanded uncertainty should include the uncertainty of the measurement process (labeled as CMC, though it is the CMC Uncertainty Component), as well as the UUT's resolution. There are some instances in which the UUT's repeatability is substituted with that of the best existing device used for calibration, as referenced in ILAC P-14: 09/2020.

Not accounting for the UUT's resolution can result in an increased risk to the end-user unless the same resolution was accounted for in the CMC uncertainty component.

The Effect of UUT Resolution on Risk & Uncertainty

One necessary contributor to the TUR denominator is the resolution of the UUT. The importance of UUT resolution to total risk is shown in Figure 7. The risk starts to increase quite dramatically as the resolution increases.

As the resolution of the device increase, so does the overall uncertainty. Figure 8 shows the relationship between resolution and Expanded Uncertainty. When the resolution is 0.001 kgf, it is insignificant. At 0.01 kgf, it is 11.52 % of the overall budget, and when raised to 0.05 kgf, it becomes dominant.

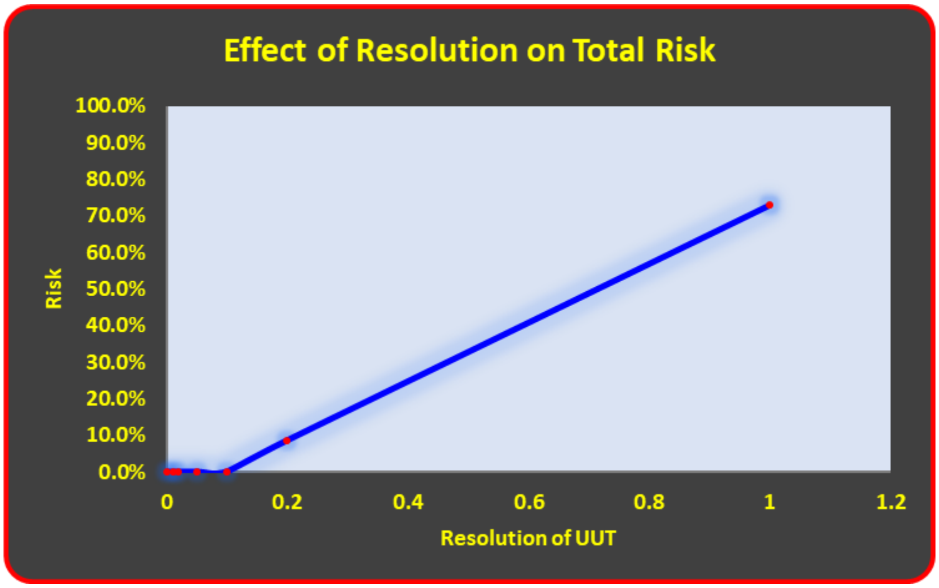
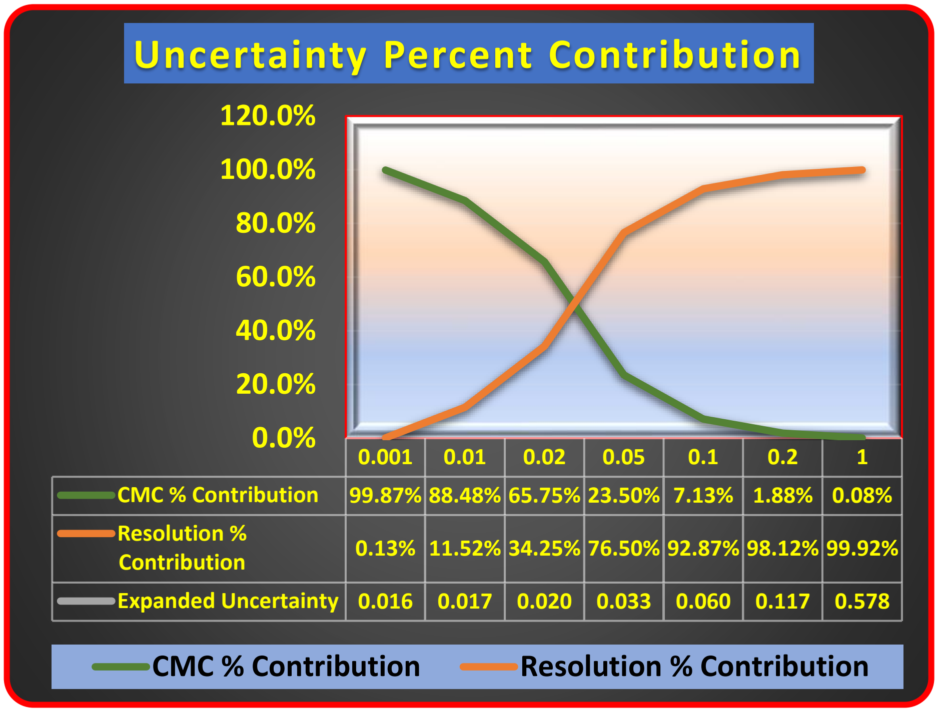


Figure 7: Resolution and the Effect on Total Risk Using a 1 000 kgf Morehouse Load Cell and Varying the Indicator Resolution (No repeatability)



	0.001	0.01	0.02	0.05	0.1	0.2	1
CMC % Contribution	99.87%	88.48%	65.75%	23.50%	7.13%	1.88%	0.08%
Resolution % Contribution	0.13%	11.52%	34.25%	76.50%	92.87%	98.12%	99.92%
Expanded Uncertainty	0.016	0.017	0.020	0.033	0.060	0.117	0.578

Figure 8: Resolution as a percentage of the Total Measurement Uncertainty Using a 1000 kgf Morehouse Load Cell and Varying the Indicator Resolution.

Figure 9: Morehouse Load Cell and Gauge Buster (UUT Example), which is used to Measure Force

Since a device with too coarse resolution will increase the measurement uncertainty, some in the Industry have created workarounds. An example of these workarounds is found in terminology such as Test Uncertainty or Test Value Uncertainty. Test Value Uncertainty was first introduced by ISO 14253-5:2015, which defines it as "measurement uncertainty associated with a test value."[14]

ISO clarifies Test Value Uncertainty by stating:

- The test value uncertainty is not a measure of the performance of the indicating measuring instrument under test; the performance is captured by the test values.
- The test value uncertainty is commonly used in the application of decision rules.
- The test value uncertainty is usually controlled by and is the responsibility of the tester, who usually provides and uses the test equipment. See 7.4 when alternative test equipment is provided by the tester counterpart (3.14).
- The test value uncertainty does not include any definitional uncertainty due to the possible non-uniqueness of test values in a permissible test instance. By agreement on the test protocol, the test is valid for any permissible test instance, for each of which a unique test measurand applies (see 3.4 Note 1 to entry).
- The test value uncertainty reveals neither the effectiveness of a test protocol in assessing a metrological characteristic, nor the reproducibility of a test value over different permissible test instances.[15]

It is important not to confuse CPU with Test Value Uncertainty. They are two different concepts. CPU and TUR calculations include contributors from the UUT that the Test Value Uncertainty does not. If adhering to the common practice of requesting a TUR > 4:1 (other guides and standards may recommend different minimum ratios) before making a statement of conformity, then the proper formula for TUR must be followed. Realizing the problem with other guides and standards, JCGM 106:2012_E states, "Care has to be taken when such rules are encountered because they are sometimes ambiguously or incompletely defined."[16]

TUR cannot be the ratio of the Manufacturer's accuracy tolerance to the reference standard uncertainty, per ANSI/NCSL Z540.3 and ILAC-G8:09/2019. Figure 10 shows a comparison of what happens to TUR when the resolution is considered and when it is not.

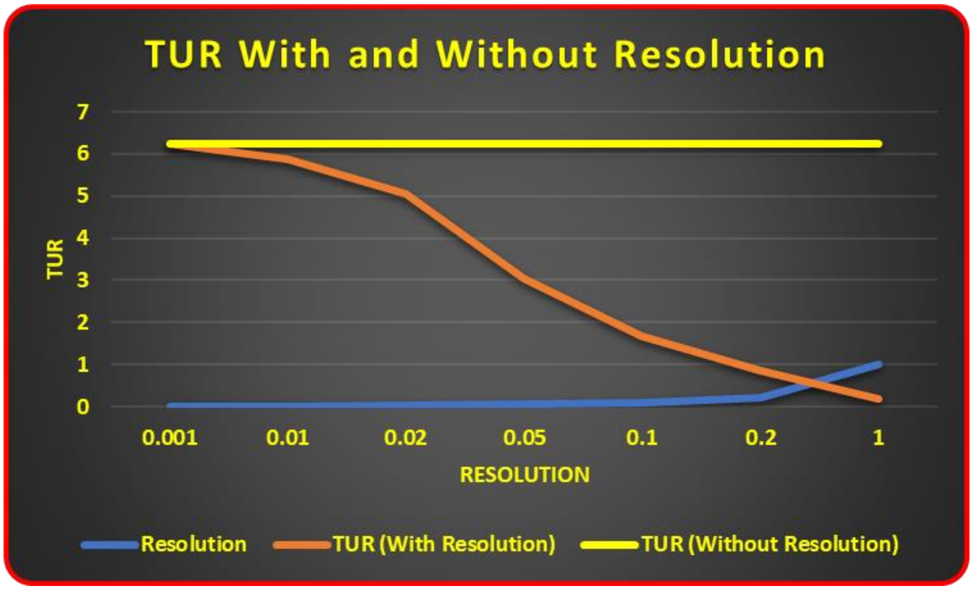


Figure 10: TUR with and without UUT Resolution

When the resolution is considered, the TUR starts at 6.25:1 with a UUT resolution of 0.001 kgf and then declines to 0.17:1 with a UUT resolution of 1.0 kgf. When the resolution is not accounted for, the TUR ratio stays at 6.25:1 regardless of the resolution. If a calibration laboratory uses the Test Value Uncertainty, then the UUT's resolution could be ignored in the conformity assessment.

Outdated Practices Lead to Higher Risk

Test Accuracy Ratio (TAR) is an outdated calculation that is not sustainable. It is the ratio of the accuracy tolerance of the unit under calibration to the accuracy tolerance of the calibration standard used.

TAR was created by Jerry Hayes and Stan Crandon in the 1950s. However, as Scott M. Mimbs points out in his paper, *Measurement Decision Risk – The Importance of Definitions*, "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."[17]

Mimbs also describes the difference between TAR and TUR in detail. He proposes that early definitions of the TUR's denominator were not well defined, which led to inconsistent applications: "The denominator for the ANSI/NCSL Z540.3 TUR is explicitly defined, thus providing better uniformity in the application of the TUR."[18] (For a Critique of 4:1 TUR Requirement, refer to NCSLI RP-18 clause 3.5.2)[19]

Many in the metrology community have invalidated TAR because it does not align with metrological traceability practices. Since the Test Value Uncertainty does not include essential uncertainty components such as the UUT resolution, it is more in unison with the outdated TAR.

As noted, TAR is not sustainable. Figure 11 shows how a 4:1 TAR works for process measurements traced back to SI units through the BIPM. In the TAR example on the left, the NMI would need a measurement process that is 256 times greater than the process measurements used in the general Industry.

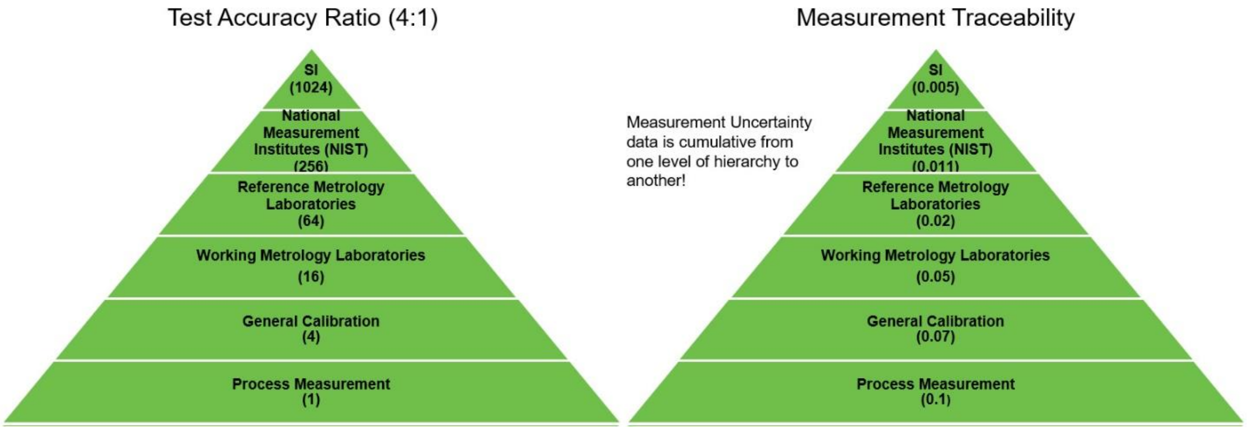


Figure 11: TAR versus TUR, illustrating how TAR is not sustainable

Mimbs provides an example of a digital micrometer using a TAR 25:1 ratio. Comparing this example with the definition of TUR found in the ANSI/NCSL Z540.3 Handbook produces a 1.5:1 ratio for the same measurement. Consequently, Mimbs concludes that computing power today is powerful enough to define risk correctly. The details are clearly defined in the ANSI/NCSL Z540.3 Handbook.

Summary

Original equipment manufacturers (OEMs) can influence standards-making committees to draft test protocols, which shift more risk to the Industry. However, does this practice comply with the ISO/IEC 17025:2017 standard and accreditation guidelines? The metrology community must recognize mandatory policy documents such as ILAC-P14 and guidance documents such as the ANSI/NCSLI Z540.3 Handbook. These documents correctly define the calibration process measurement uncertainty used for calibration.

If a device's uncertainty has too coarse of a resolution, then a device that does not subject the conformance decision to higher risk is needed. Furthermore, if uncertainty contributors are omitted at the OEM level of calibration, how will we predict the global consumer's risk and adequately account for uncertainty in the measurement traceability chain?

This paper has presented several examples to demonstrate why best practices should be followed for conformity assessments. These best practices are outlined in guidance documents, policy documents, and standards documents such as ANSI/NCSLI Z540.3, ILAC-P14, and ISO/IEC 17025:2017.

Consider manufacturers and calibration laboratories that fail to correctly calculate both uncertainty and risk on equipment used to test medical equipment, airplanes, cars, and bridges. How will you feel when you schedule your next surgery, sit in a traffic jam on a bridge, or experience mechanical problems on your next flight?

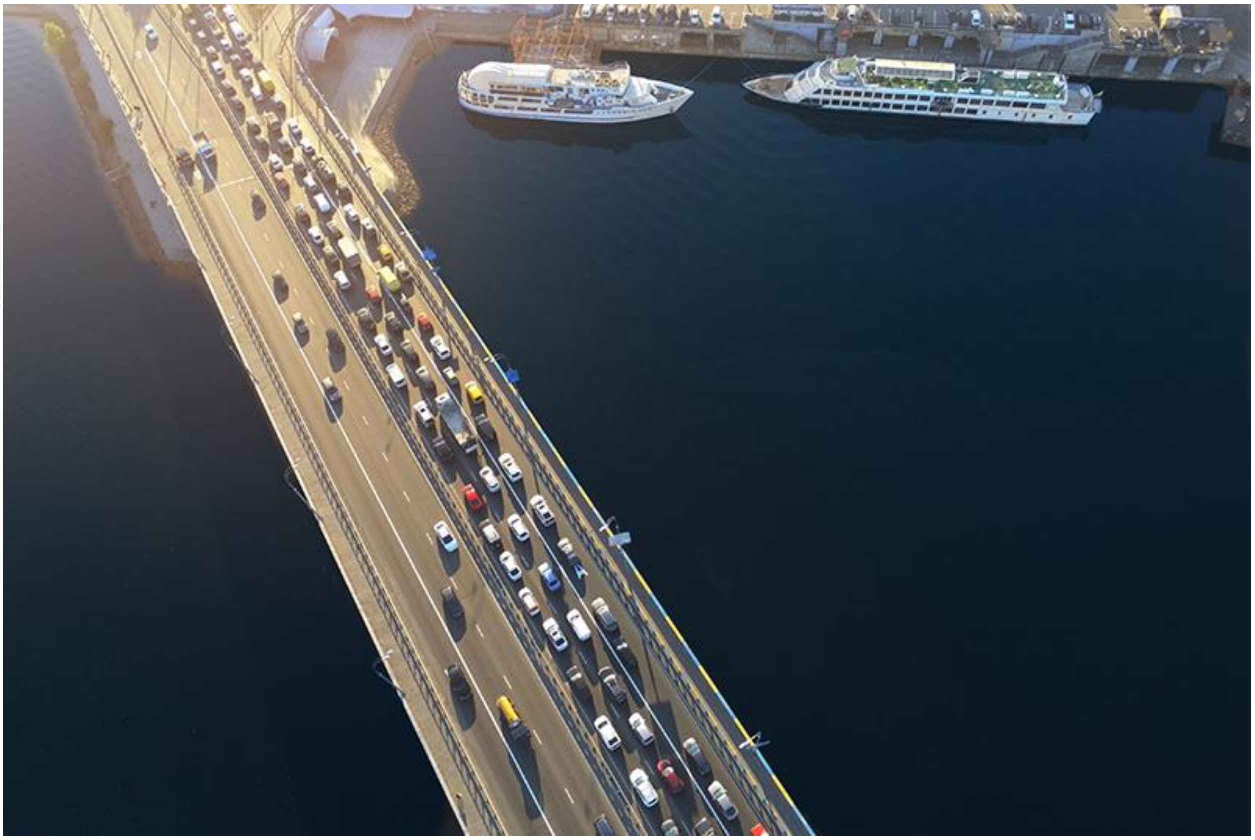


Figure 12: Risk Considerations in a Traffic Jam on a Bridge

Will you feel confident that measurements were performed correctly and, when applied, will keep you safe—or will you worry about your safety? How do you feel now knowing that someone may have failed to calculate and apply measurement decision risk correctly? Having to question if the people making the measurements may have understated the calibration measurement process uncertainty does not boost confidence.

---

[1] ANSI/NCSLI Z540.3-2006 "Requirements for the Calibration of Measuring and Test Equipment."
[2] ILAC G8:2019 "Guidelines on Decision Rules and Statements of Conformity."
[3] *To Engineer is Human: The Role of Failure in Successful Design*, by Henry Petroski.
[4] JCGM 106:2012_E clause 3.3.15 "Evaluation of measurement data – The role of measurement uncertainty in conformity assessment."

---

[5] JCGM 200:2012 International Vocabulary of Metrology (VIM).

[6] JCGM 200:2012 International Vocabulary of Metrology (VIM).

[7] ISO/IEC 17025:2017 "General requirements for the competence of testing and calibration laboratories," clause 7.8.6.1

[8] UKAS LAB 48 Decision Rules and Statements of Conformity

[9] ISO/IEC 17025:2017 "General requirements for the competence of testing and calibration laboratories," clause 7.8.6.1

[10] ANSI/NCSL Z540.3 Handbook "Handbook for the Application of ANSI/NCSLI 540.3-2006 - Requirements for the Calibration of Measuring and Test Equipment."

[11] ANSI/NCSL Z540.3 Handbook "Handbook for the Application of ANSI/NCSLI 540.3-2006 - Requirements for the Calibration of Measuring and Test Equipment."

[12] ILAC P-14:09/2020, "Policy for Uncertainty in Calibration," clause 5.4

[13] ILAC P-14:09/2020, "Policy for Uncertainty in Calibration," clause 4.3 note 2

[14] ISO 14253-5:2015

[15] ISO 14253-5:2015

[16] JCGM 106:2012_E clause 3.3.15 "Evaluation of measurement data – The role of measurement uncertainty in conformity assessment."

[17] Measurement Decision Risk – The Importance of Definitions, by Scott M. Mimbs

[18] Measurement Decision Risk – The Importance of Definitions, by Scott M. Mimbs

[19] NCSLI RP-18 – Estimation and Evaluation of Measurement Decision Risk