17th IMEKO TC 10 and EUROLAB Virtual Conference
"Global Trends in Testing, Diagnostics & Inspection for 2030"
October 20-22, 2020.

Quantification of Misunderstanding: A Case of Tolerance Limits against Calibration Uncertainty in ISO/IEC 17025 Accredited Laboratories

Filip Flego[1], Marija Pavlović[2], Siniša Prugovečki[3]

[1]*Metroteka, Zagreb, Croatia, filip.flego@metroteka.com*
[2]*Metroteka, Zagreb, Croatia, marija.pavlovic@metroteka.com*
[3]*Metroteka, Zagreb, Croatia, sinisa.prugovecki@metroteka.com*

***Abstract* – Measurement uncertainty undoubtedly plays a large role in the assessment of compliance with certain specifications. For example, the widely accepted ILAC G8:09/19 publication gives relatively flexible guidelines for assessing the conformity of certain instruments to their respective standards or specifications. On the other hand, the information given in calibration certificates is often interpreted too rigorously which can, unfortunately, lead to the false conclusion that the instrument in question doesn't actually meet the client's specified requirements.**

In the following paper we will examine the role measurement uncertainty has in the assessment of compliance with regards to specified tolerance limits for 5 different instruments: calipers, contact thermometers (at temperatures above 600 °C), hygrometers, temperature chambers (such as laboratory incubators or refrigerators) and piston pipettes.

Keywords – Measurement uncertainty, Calibration, Compliance, Tolerance limits.

I. INTRODUCTION

Although it is understood to be a relatively simple mathematical concept, measurement uncertainty often contains a certain dose of ambiguity within it, which can then further lead to its misunderstanding, misrepresentation and sometimes even falsification. However, this is mainly due to misguidance within the metrology community, and not necessarily due to the users of the instruments in question who often times don't fully understand the information contained in the calibration certificate handed to them and how to properly interpret the results.

In this paper we will attempt to contribute to a better general understanding of measurement uncertainty within the context of assessing whether an instrument conforms to requirements regarding its specified tolerance limits, while also complying to the ILAC-G8:09/2019 Guidelines on Decision Rules and Statements of Conformity. By analysing raw data from 5 different types of instruments, we will compare the percentage of instruments which successfully conform to their respective standards or specifications taking measurement uncertainty into account, with the percentage of instruments which successfully conform without taking measurement uncertainty into account. We will then try to put those numbers into a valuable and revealing perspective.

II. COMMON PRACTICE IN CONFORMITY ASSESSMENT

Generally, the users of instruments who, upon receiving their calibration certificates, assess whether they meet specified requirements or not, do this in one of two ways: either they *do* take the whole expanded (k=2) measurement uncertainty stated in the instrument's calibration certificate into consideration and then add it to the instrument's error (or some other parameter unfortunately), or they *do not* consider it at all.

At first glance, it seems perfectly reasonable to necessarily take measurement uncertainty into account. After all, this is probably the most common practice in conformity assessment. However, as we will later attempt to show, and as ILAC-G8:09/2019 states, there are more than these two options available to us and it is not uncommon to assess whether an instrument lies within its tolerance limits by purposefully not taking measurement uncertainty into account.

With respect to accredited testing or calibration laboratories, maintaining compliancy with ISO/IEC 17025:2017 is mandatory, so it is important not to have clashing demands within the various standards these laboratories have to conform to. Having said that, within

ISO/IEC 17025:2017 there is no specific requirement regarding the consideration of measurement uncertainty in conformity assessment, which actually prevents possible inconsistencies with requirements from other documents. Instead, the only practical requirement is evaluating measurement uncertainty and carefully documenting the decision rule upon which conformity is based; also, consistently abiding by this internal process at all times, with the main focus being on risk analysis [2].

III. CORRECTLY ASSESSING CONFORMITY

As previously mentioned, the ILAC-G8:09/2019 Guidelines on Decision Rules and Statements of Conformity is commonly considered to be the go-to standard for conformity assessment within accredited laboratories. When considering measurement uncertainty, it clearly states: *"If measurement uncertainty is taken directly into account, the acceptance interval will be a restricted part of the tolerance. The larger the measurement uncertainty is, the smaller the acceptance interval gets. This will result in fewer accepted results than if measurement uncertainty had been smaller [1]"*. This means that the decision rules upon which conformity assessment is based on can not only be very different depending on the instruments in question, but can also quickly become complicated to define.

The following figure (Fig. 1) is a graphic representation of binary acceptance criteria; the dot represents an instrument's error and the error bars represent its expanded (k=2) measurement uncertainty [1]. Depending on where each error lies within certain specified tolerance limits, it should be apparent whether the measured value does or doesn't meet its requirements.

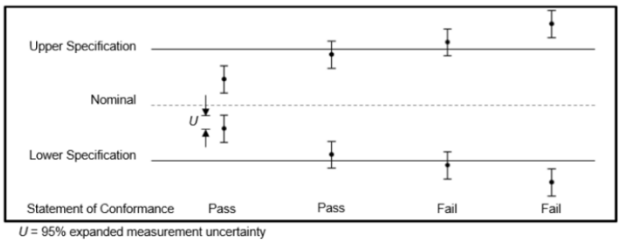

Fig. 1. A binary statement regarding acceptance criteria with reference to expanded (k=2) measurement uncertainty

By only considering where the error lies and ignoring the measured value's uncertainty, one can conclude that the first two cases give an acceptable value – if the measured value is within the acceptance limit (AL), it passes, and if it is out of the acceptance limit, it fails. However, by taking measurement uncertainty into account, and stating that it too must lie within the defined acceptance limits in order for a measured value to pass, all cases fail except the first, meaning it is less likely an instrument will meet its specified requirements.

By introducing a guard band (w), the probability of making an incorrect conformance decision decreases. This is because, in effect, it reduces the acceptance limit below that of the tolerance limit (TL), thus functioning as a safety factor built into the measurement decision process. This is often done to account for measurement uncertainty. The difference between these two values is the actual length of the guard band (w = AL – TL) [1].

As shown in the following figure (Fig. 2), it is noticeable that acceptance is determined by the measuring result being within the acceptance limits. If the measurement error doesn't lie within the acceptance limits, it fails to meet the defined requirements.

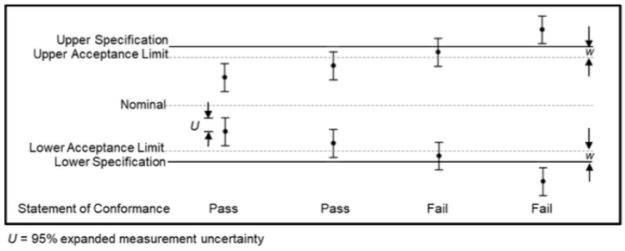

Fig. 2. A binary statement regarding acceptance criteria with a guard band (w)

Therefore, it is important to point out that the length of the guard band can determine whether an error is acceptable or not. This means that conformity with a requirement is inherently connected to the decision rule employed, which furthermore means that the decision rule should be agreed upon before the measurements are taken.

IV. CONSIDERATION OF MEASUREMENT UNCERTAINTY

Simply put, measurement uncertainty can be characterised as "quantified doubt" in the process of measuring something. With regards to conformity assessment, if expanded (k=2) measurement uncertainty (the *k* factor reflecting the number of standard deviations used, resulting with a confidence level of approximately 95% in this case) is taken directly into account by adding it to the error, the acceptance interval of a result will be a reduced part of the tolerance interval previously described. The larger the measurement uncertainty is, the smaller the acceptance interval gets, which will therefore result in fewer measurements being accepted than if it had been smaller to begin with.

Practically speaking, it is the instrument's users who usually define the acceptable tolerance limits for the particular instrument in question. This is normally based on compliance to specific standards, recommendations, their own experience using the instrument, or lastly, the manufacturer's specifications. With the latter too often being unrealistic, unattainable and untrue, standards and recommendations generally contain well defined

tolerance limits, however they tend to be out-dated because they often don't consider measurement uncertainty. On the other hand, when the instrument's users define the acceptable tolerance limits, it is common practice to vastly undervalue it. This is obviously wrong as it is extremely important, and regularly overlooked, that measurement uncertainty is first and foremost objectively and realistically calculated and portrayed, not falsified.

V. DATA ANALYSIS

As mentioned, the goal of this paper is to better understand how and whether or not to consider measurement uncertainty in the process of determining an instrument's conformity to a specified standard, with regards to that instruments' permissible tolerance limits.

For this purpose, the data gathered from 5 different types of instruments was analysed in order to gain insight into the important differences between considering and not considering measurement uncertainty. The data was collected from years worth of calibrating these instruments, resulting in a total of hundreds of calibration certificates from each instrument type analysed. From this data, for each type, the objective was to determine the percentage of instruments which successfully conform to their specified requirements.

As for the measurement uncertainties depicted in each calibration certificate, taken into account for each of the type of instrument analysed, it is safe to say that they are in the typical range of values within the calibration laboratory field. All these certificates originate from Metroteka, a private calibration laboratory based in Croatia.

A. Calipers

Calipers (both Vernier and digital) are a simple and well-known instrument for length measurement, and are one of the most common instruments brought in calibration laboratories all over the world. For the purpose of this paper, the errors of 446 callipers were analysed; all of them having the same span of 150 mm in order to preserve consistency, and differing only by their resolution. The errors were then compared to the permissible tolerance limits defined in the internationally acclaimed German standard – DIN 862 Geometrical product specifications (GPS) – Callipers – Maximum permissible errors.

According to the latest version of DIN 862, in order to meet the necessary requirements, that is to say, conform to the specified tolerance limits, the margin of error for callipers with a resolution of 0,01 mm or 0,02 mm must not exceed 0,025 mm (obtained by interpolation, based on the calliper's span) or, for a resolution of 0,05 mm, it must not exceed a margin error of 0,05 mm, as shown in Table 1 [3].

Table 1. Maximum permissible errors [µm] according to DIN 862.

Span [mm]	Resolution [mm]		
	0,01	0,02	0,05
50	20	20	50
100	20	20	50
200	30	30	50

The data shows the following: out of exactly 446 calibrated callipers, based upon the mentioned requirements from DIN 862, approximately 75% of them (334 out of 446) meet those requirements when measurement uncertainty is *not* taken into account. However, interestingly, when uncertainty *is* taken into account, the number of calipers which successfully conform to exactly the same requirements falls drastically – just under 1% of them (4 out of 446) successfully conform to the specified requirements.

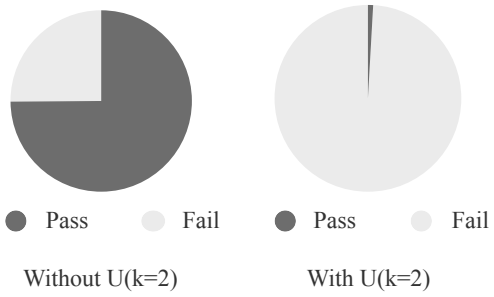

Fig. 3. Percentage of acceptable calipers

This surprising result is due to the fact that it is extremely uncommon for a calibration laboratory to be able to realistically express its measuring uncertainty for calipers below 0,025 mm.

B. Contact thermometers – type K thermocouples

A thermocouple is an electrical device consisting of two dissimilar electrical conductors which produce a temperature-dependant voltage; this voltage can then be interpreted to measure temperature. Thermocouples can measure a wide range of temperatures, however they tend not to be extremely accurate, especially at higher temperatures, so a thermocouple error of less than 1 ºC is rarely achieved.

For the purpose of this paper, only calibration results at temperatures above +600 ºC will be considered, which presupposes the use of *type K* thermocouples, as they are most commonly used for such temperatures. According to the latest version of IEC 60584-1, in order to meet the specified requirements, the margin of error for such thermocouples must not exceed 1,5 ºC for class 1, or 2,5 ºC for class 2 type k thermocouples [4], shown in Table 2.

Table 2. Tolerance values by class for type-K thermocouples according to IEC 60584-1

Class	Temperature range	Tolerance value
1	-40 ... +1000 °C	±1.5 °C or 0.0040 · |t|
2	-40 ... +1200 °C	±2.5 °C or 0.0075 · |t|
ItI *is the absolute value of the temperature measured in °C*		

After analysing 314 calibrated type K thermocouples, the data shows the following: for temperatures above +600 ºC, around 54% (169 out of 314) of thermocouples meet the specified requirements for class 1 *without* measurement uncertainty taken into account. On the other hand, for class 2 thermocouples, approximately 81% (255 out of 314) of them meet the specified requirements, also *without* considering measurement uncertainty. However, when measurement uncertainty *is* taken into account, the amount of thermocouples which successfully conform to the exact same requirements, once again, falls drastically – exactly 0 for class 1 requirements and 19% (59 out of 314) for class 2 requirements.

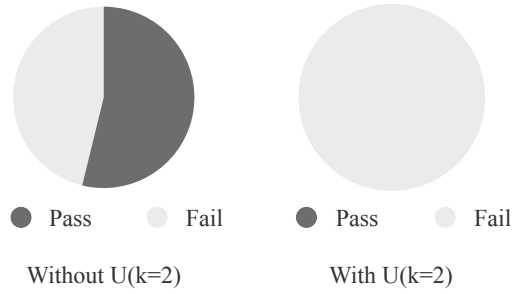

Fig. 4. Percentage of acceptable Class 1 thermocouples

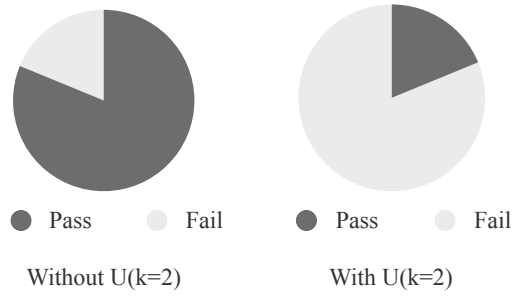

Fig. 5. Percentage of acceptable Class 2 thermocouples

C. Hygrometers

Hygrometers are instruments used to measure the amount of water vapor in a medium, most commonly in air. Such instruments usually respond to, or are calibrated to read relative humidity (RH), which refers to the moisture content (i.e., water vapor) of the atmosphere, expressed as a percentage of the amount of moisture that can be retained by the atmosphere (moisture-holding capacity) at a given temperature and pressure without condensation. In other words, a reading of 100% RH means that the air is totally saturated with water vapor.

The type of hygrometers analysed in this paper are the most common type used in the industry – capacitive hygrometers which measure the effect of humidity on the dielectric constant of the material they are made of.

There are no commonly accepted standards for which conformity is mandatory for hygrometers, which can possibly be challenging for instruments users while interpreting the calibration results. So, for the sake of this article, since the manufacturers' specifications for hygrometers are very often untrue, we will consider the accepted accuracy for hygrometers is ± 5% RH in the range of 5% to 95% RH, which is most often used in conformity assessment in laboratories, especially within the rigorous pharmaceutical industry.

The errors taken from the calibration certificates of 2711 hygrometers were analysed so as to gain insight in the role measurement uncertainty has in assessing whether the errors fall within the defined tolerance limits of ± 5% RH. The data shows the following: around 85% of all analysed calibrated hygrometers (2300 out of 2711) meet the defined requirements *without* consideration of measurement uncertainty. When measurement uncertainty *is* considered, the percentage of hygrometers which conform to the same requirements falls to around 53% (1440 out of 2711).

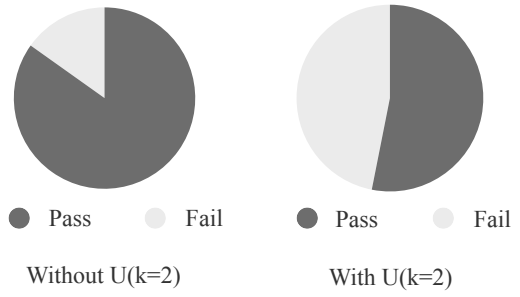

Fig. 6. Percentage of acceptable hygrometers

D. Piston pipettes

Piston pipettes are laboratory instruments used to transport a measured volume of liquid from one container to another. As they are a vital piece of equipment in biology, chemistry and especially the medical industry, these measured volumes are often extremely small and their accuracy varies greatly depending on the type of pipette in question.

The type considered in this paper are piston-driven pipettes which are a type of micropipette; namely, they handle volumes of liquid in the microliter scale.

With the objective of this paper in mind, the errors of 3127 such pipettes were taken from their calibration certificates and were then compared to the permissible tolerance limits for maximum permissible systematic error and maximum permissible coefficient of variation defined in the standard ISO 8665-2, as shown in Table 3 and Table 4 [5].

Table 3. Maximum permissible systematic errors according to ISO 8655-2

Nominal volume	Max. permissible systematic error	
[µl]	± %	± [µl]
10	1,2	0,12
20	1,0	0,2
50	1,0	0,5
100	0,8	0,8
200	0,8	1,6
500	0,8	4,0
1000	0,8	8,0
2000	0,8	16
5000	0,8	40
10000	0,6	60

Table 4. Maximum permissible coefficient of variation according to ISO 8655-2

Nominal volume	Max. permissible coeff. of variation	
[µl]	± %	± [µl]
10	0,8	0,08
20	0,5	0,1
50	0,4	0,2
100	0,3	0,3
200	0,3	0,6
500	0,3	1,5
1000	0,3	3
2000	0,3	6
5000	0,3	15
10000	0,3	30

When assessing whether piston-pipettes meet their specified requirements with regards to the maximum permissible errors stated in the previous tables, it is important to point out that they must necessarily meet two separate criteria. The first has to do with the pipette's relative systematic error which is calculated somewhat unorthodoxly compared to other instruments because it *always* depends on the pipette's maximum volume, not the set volume for the particular measurement being carried out. The other criterium has to do with the pipette's coefficient of variation which is directly connected with the standard deviation calculated from 10 measurements.

With this in mind, analysing the results from the pipette's calibration certificates showed the following: approximately 94% of them (3008 out of 3217) have errors which lie within the specified values regarding both systematic error and coefficient of variation *without* the consideration of measurement uncertainty. Furthermore, around 94% (3023 out of 3217) meet the requirements for maximum permissible systematic errors (Table 3.), while around 96% (3077 out of 3217) meet the requirements for maximum coefficient of variation (Table 4.); both *without* taking measurement uncertainty into account.

Considering measurement uncertainty we, however, get the following results: both criteria are met amongst 45% (1460 out of 3217) of the piston-pipettes analysed. Additionally, around 76% (2428 put of 3217) piston-pipettes meet the requirements for maximum systematic error, while only 48% (1528 out of 3217) meet the requirements for the coefficient of variation.

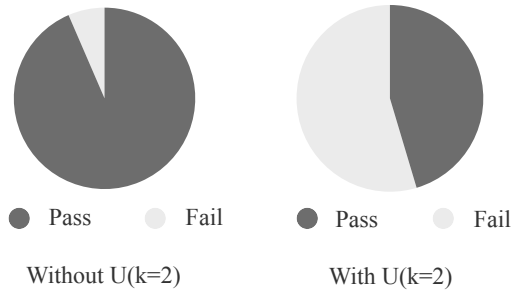

Fig. 7. Percentage of acceptable piston-pipettes considering both systematic error and coefficient of variation

Due to the specificity of piston-pipettes, it is unavoidable to note the following: to take measurement uncertainty into account with regards to the piston-pipette's coefficient of variation is a nonsensical act. The reason being that the CV is calculated as the standard deviation of a tenfold measurement [5]. Moreover, taking the standard deviation of measurement into consideration is *mandatory* for calculating the expanded (k=2) measurement uncertainty of any given instrument. Therefore, it is unnecessary and unreasonable to consider the measurement uncertainty of the standard deviation of an instrument's measurements when assessing that instrument's conformity to a standard. However, it is worth noting that we are aware of cases where this was requested from an ISO/IEC 17025 accredited calibration laboratory during its assessment performed by a national accreditation agency. This is another great example of how measurement uncertainty is vastly misunderstood and abused within the metrology community.

E. Temperature chambers

Although temperature chambers (freezers, refrigerators, incubators, dry sterilizers etc.) are not actual measuring instruments, rather temperature-controlled enclosures used in all kinds of testing activities, as such they undergo the same process of conformity assessment as actual measuring instruments.

Unfortunately, in reality, apart from the problem of deciding whether or not and how to use measurement uncertainty in conformity assessment of temperature chambers, ISO and other published standards for test methods almost never actually define the tolerance limits, although it seems they do. For example, when a standard says that the temperature in a particular temperature chamber should be (+37 ± 1) °C, it is not clear whether this criterium should be applied to the average temperature in the reference location, or to the average temperature at all locations, or to the temperature at all locations at all times, or finally - to the temperature of all types of material samples at all locations at all times. This is a huge problem, but it is outside the scope of this paper, so for the sake of our analysis we shall show the results in 3 cases:

a) the acceptance criteria has to be met for average temperatures at all locations (*spatial homogeneity** should be less than the tolerance limit)
b) the acceptance criteria has to be met for temperatures at all locations at all times (the sum of spatial homogeneity and *temporal stability** should be less than the tolerance limit)
c) the acceptance criteria has to be met for temperatures at all locations, at all times, and for every type of material (the sum of spatial homogeneity, temporal stability and the *influence of thermal radiation** should be less than the tolerance limit)

** Note: These parameters are defined according to DAkkS-DKD-R 5-7 (Calibration of climatic chambers) [6].*

Additionally, all 3 parameters, as defined in DAkkS-DKD-R 5-7, have to be included in the calculation of the expanded (k=2) measurement uncertainty. So, once again, similarly to the coefficient of variation previously mentioned, it is nonsensical to add measurement uncertainty to them, but nevertheless this is in fact common practice. We will also assume that the deviation of temperature in the reference location in the temperature chamber is corrected after the calibration results are issued to the user, so the only 3 parameters which can contribute to the conformity assessment are those already mentioned, with or without measurement uncertainty, of course.

For the sake of this article, as tolerance limits we will consider the most frequently used limits in testing laboratories: ±5 °C for freezers (set point below 0 °C), ±3 °C for refrigerators (set point between 0 °C and +15 °C), ±1 °C for incubators (set point between +15 °C and +50 °C), ±2 °C for chambers with a set point between +50 °C and +100 °C and ±5 °C for dry ovens and sterilizers (set point between +100 °C and +250 °C), shown in Table 5.

Table 5. Maximum permissible errors [°C] for temperature chambers

Temperature chamber	Tolerance limt [°C]
Freezer	± 5
Refrigerator	± 3
Incubator	± 1
Chamber	± 2
Dry oven	± 5
Sterilizer	± 5

Overall, the results from 618 temperature chambers were considered. Analysing the results *without* taking measurement uncertainty into account showed the following: approximately 76% of chambers (468 out of 618) meet the criteria for average temperatures in all locations, 52% of chambers (319 out of 618) meet the criteria for temperatures in all locations at all times, and 44% of chambers (271 out of 618) meet the criteria for temperatures in all locations at all times and for all types of sample materials.

Taking measurement uncertainty into account (despite it being senseless) showed the following: approximately 16% of chambers (100 out of 618) meet the criteria for average temperatures in all locations, 11% of chambers (69 out of 618) meet the criteria for temperatures in all locations at all times, and just 8% of chambers (49 out of 618) meet the criteria for temperatures in all locations at all times and for all types of sample materials.

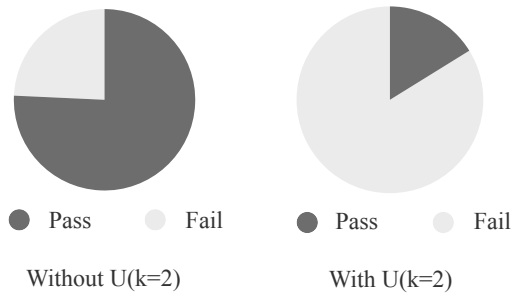

Fig. 7. Percentage of acceptable temperature chambers considering spatial homogeneity

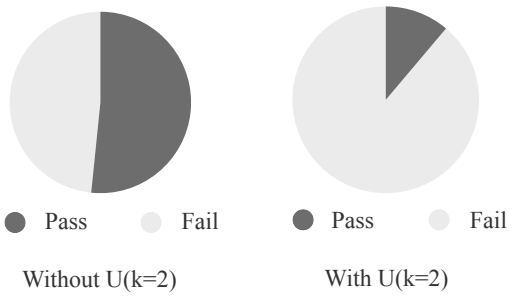

Fig. 8. Percentage of acceptable temperature chambers considering spatial homogeneity and temporal stability

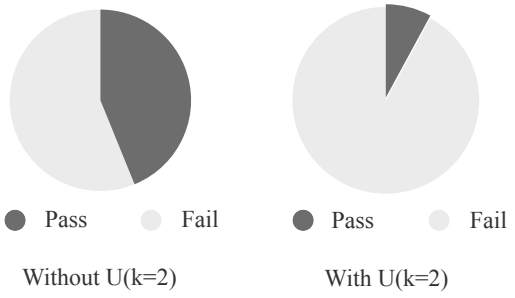

Fig. 9. Percentage of acceptable temperature chambers considering spatial homogeneity, temporal stability and the influence of thermal radiation

VI. CONCLUSIONS AND OUTLOOK

To conclude, when assessing whether an instrument's errors, based on its calibration certificates, fall within specified tolerance limits (or criteria), despite often contradictory recommendations, the data supports the notion that it is almost always better *not* to take measurement uncertainty into account, because otherwise a very small percentage of measuring instruments used would meet the criteria specified. This is due to the fact that, at the time tolerance limits in standards or internal requirements for instruments in companies were defined, measurement uncertainty was not thought about at all, let alone considered in conformity assessment. Furthermore, as we have stated earlier, it is still vastly misunderstood and far too often expected to be unrealistically small. In fact, it is not at all uncommon for certain newer standards to explicitly remark that measurement uncertainty is not to be considered during conformity assessment, as in, for example, ISO 6789:2017 for torque wrenches [7].

For this reason, the most sensible thing to do is to follow the very liberal and loose recommendations stated in ILAC-G8:09/2019, while also, when it comes to calibration or testing laboratories, fully complying to ISO/IEC 17025. This means establishing an internal procedure to follow, in which the consideration of measurement uncertainty with regards to conformity assessment is accurately described and defined, and, most importantly – following this procedure precisely and consistently.

REFERENCES

1. ILAC-G8:09/2019 *Guidelines on Decision Rules and Statements of Conformity,* 2019
2. ISO/IEC 17025:2017 *General requirements for the competency of testing and calibration laboratories,* 2017
3. DIN 862:2015-03 *Geometrical product specifications (GPS) – Callipers – Maximum permissible errors,* 2015
4. IEC 60584-1:2013 *Thermocouples - Part 1: EMF specifications and tolerances,* 2013
5. ISO 8655-2:2002 *Piston-operated volumetric apparatus – Part 2: Piston pipettes,* 2012
6. DAkkS-DKD-R 5-7 *Calibration of Climatic Chambers,* 2010
7. ISO 6789:2017 *Assembly tools for screws and nuts - Hand torque tools*