UNCERTAINTY

# Uncertainty of the calibrating instrument, confidence in the measurement process and the relation between them

TADEUSZ SKWIRCZYŃSKI
Independent Consultant, Warsaw, Poland

*Abstract*

*For a given error distribution, confidence in the measurement process depends on the test uncertainty ratio (TUR) and on the confidence interval. When selecting a measuring instrument or measurement standard to carry out a calibration or verification or, in general, a measurement, this dependence becomes a vital issue.*

*The author has considered the effect of several TURs encountered in practical situations on incorrect test decisions. This consideration has also been extended to the effect on correct test decisions, reliability of test results and confidence in the measurement process for normal error distribution, for both the equipment under test (EUT) and the calibrating instrument, at two confidence interval specifications.*

*This paper contains a short presentation of specific relevant definitions and issues, results of the study and discussion, and two examples of a lack of specific information on the TUR in certain standards. The analysis has been performed for TURs ranging from 1:1 to 100:1 and for confidence interval specifications of 2σ and 3σ.*

*Both the information given and the conclusions which have been drawn can be used in calibration and verification and, generally, also in measurement.*

## 1 Introduction

Measurements and the calibration of measuring instruments are essential aspects of activities such as maintaining quality control and quality assurance in production, complying with and enforcing laws and regulations, conducting research and development in science and engineering and calibrating and verifying measurement standards and instruments in order to achieve traceability to national standards.

Calibration is the determination, by measurement and comparison with a measurement standard, of the correct value of a reading on a measuring instrument. The calibration system considered in this paper is shown in Fig. 1. The calibrator (standard) is the source of the standard signal, and the standard value of the calibrator is compared with the measurement result indicated by the EUT.

Verification is an activity performed by a national measurement service in which similar measurement procedures are used as for calibration.

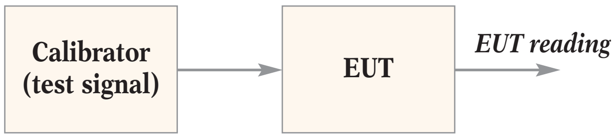



Fig. 1 Measurement system used in calibration

The overall measurement error consists of two components: the error arising in the EUT and that originating from the measurement standard [1]. It is worth mentioning that good measurement has its origins as much in the study of errors or uncertainties of the measurement as it does from the choice of the principle of measurement [2]. When reporting the result of a measurement of a physical quantity, it is therefore also necessary to state the relevant error or uncertainty of the measurement.

Uncertainty of measurement is a parameter associated with the result of a measurement that characterizes the dispersion of the values that would reasonably be attributed to the measurand [3].

Figure 2 illustrates the meaning of uncertainty and error of measurement using the normal distribution curve and shows a situation where the confidence interval ranges from – 2σ to + 2σ which corresponds to an uncertainty of 2σ at about 95.45 % confidence level, where σ is the standard deviation. In metrological practice the confidence interval is usually assumed to be from – 2σ to + 2σ or from – 3σ to + 3σ [3, 4]. In Fig. 2, the true value is – 1σ and the EUT reading is 0, so the error is + 1σ.

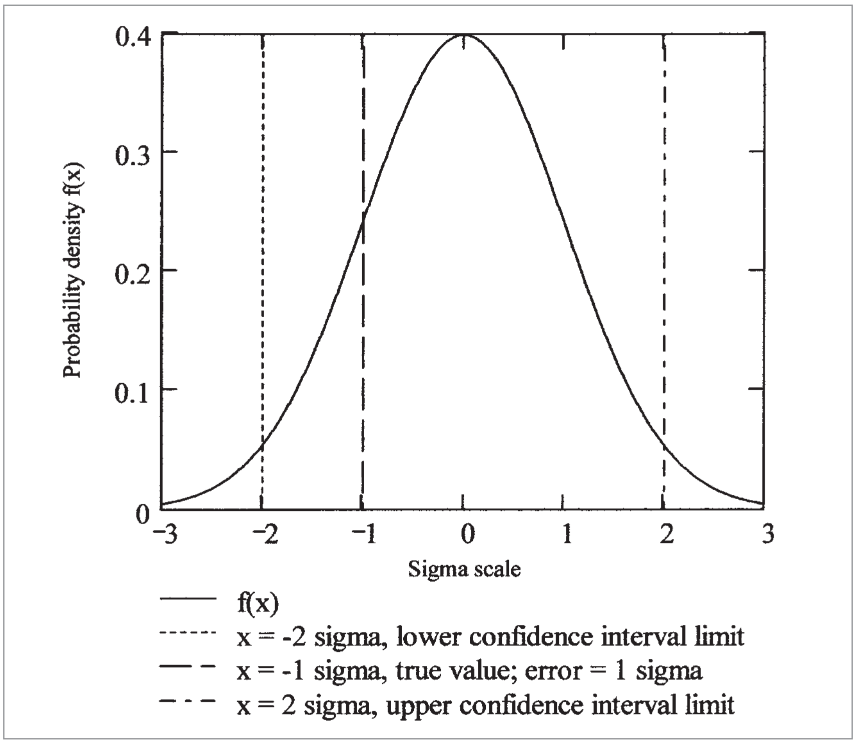



Fig. 2 Error and uncertainty

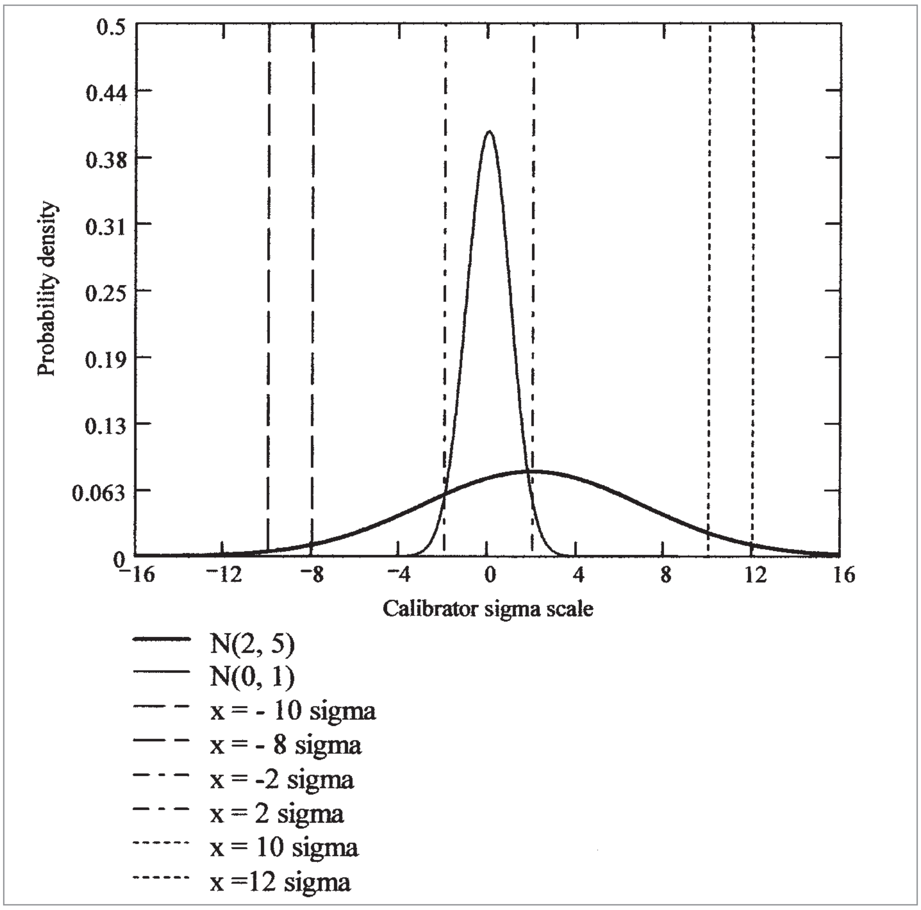



Fig. 3 Illustration of incorrect test decisions for 2σ specifications for calibrator and EUT and normal distribution of errors in their populations

## 2 Test uncertainty ratio (TUR)

The TUR for a measurand is defined as the standard uncertainty of the EUT divided by that of the calibrating instrument (measurement standard) used to test it [4, 5, 6]. A reliable TUR is only obtained when the specifications for the EUT and the calibrating instrument are correlated according to their error distributions and confidence intervals. It can be said that a reliable TUR is a *sine qua non* condition for good quality calibration. The purpose of calibration is to gain confidence that the EUT is capable of making measurements within the specifications. And, generally, the purpose of measurement is to gain confidence that the value of the measurand is within its tolerance limits. Testing laboratories need to use measuring instruments that have uncertainty specifications which are adequate for the measurements they perform.

## 3 Incorrect test decisions and confidence in the measurement process

Actual measuring instrument test results can contain four kinds of test decisions:

- acceptance of good units,
- rejection of bad units,
- rejection of good units, and
- acceptance of bad units.

The "ideal" situation is that the results consist of only the first two kinds of test decisions, the second two being the results of uncertainty in the specifications for both the EUT and the calibrating instrument.

An accepted "good" unit is a calibrated instrument that is within its specified tolerance limits and a rejected "bad" unit is one that is outside its tolerance limits. Thus, the actual test results contain correct and incorrect test decisions. Correct test decisions contain acceptance of good units and rejection of bad units whereas incorrect test decisions contain rejection of good units (incorrect "fail") and acceptance of bad units (incorrect "pass"). This situation is shown in Fig. 3 for a 5:1 TUR, normal error distribution and $2\sigma$ specifications for both the EUT and the measuring instrument (calibrator). In this example, the normal distribution curve N (0, 1) - where 0 is the mean value and 1 is the standard deviation - illustrates the error distribution for the calibrator and the normal distribution curve N (2, 5) shows the error distribution for the EUT.

As illustrated, the actual output of the calibrator is larger than the nominal output by the maximum permissible error, i.e. by $+ 2\sigma$. Relative to the EUT specification, the calibrator output is at $+ 0.4\sigma$. In terms of the test limits, the EUT readings which are truly within the tolerance limits are in the range from $- 1.6\sigma$ to $+ 2.4\sigma$. This is due to the fact that the readings have a normal distribution and so they are symmetrically distributed on either side of a stimulus that is displaced by $+ 0.4\sigma$ from its nominal value. That is why the EUT readings between $+ 2\sigma$ and $+ 2.4\sigma$ will be incorrectly outside the tolerance limits and the readings between $- 2\sigma$ and $- 1.6\sigma$ will be incorrectly within them. As the distribution of errors is normal, the number of EUT units within the tolerance limits that are incorrectly rejected exceeds the number of EUT units which are outside the tolerance limits that are incorrectly accepted.

Furthermore, as the error distribution is normal so the curve is symmetrical, and analogous results of the analysis will be obtained when the output of the calibrator is displaced to $- 2\sigma$, i.e. to $- 0.4\sigma$ relative to the EUT specification.

The decimal fraction of correct test decisions equals 1 minus the decimal fraction of incorrect test decisions (incorrect "fail" plus incorrect "pass"). The larger is the fraction of correct test decisions, the larger will be the confidence in the measurement process. It is generally assumed that 100 % correct test decisions is unattainable at any cost. On the other hand, there is usually a target value for the correct test decision percentage. This percentage depends on the activity supported by the testing. The percentage of correct test decisions below the target value will significantly decrease reliability of test results and confidence in the measurement process and may be assumed to have unacceptable effects on such factors supported by the test as human health, safety and lives, and cost of manufacturing or quality of product, to mention just a few of them.

## 4 Results of analysis and discussion

The incorrect test decisions have been studied as a function of the TUR value ranging from 1:1 to 100:1 at $2\sigma$ and $3\sigma$ confidence intervals and normal error distribution for both the EUT and the calibrator.

The results of the study are given in the form of graphs in Figs. 4–7. The graphs contain the error of the calibrator in standard deviations, as an independent variable, and the following decimal fractions of the EUT population as dependant variables:

- good units rejected (incorrect "fail" units) in Figs. 4 and 6, and
- bad units accepted (incorrect "pass" units) in Figs. 5 and 7.

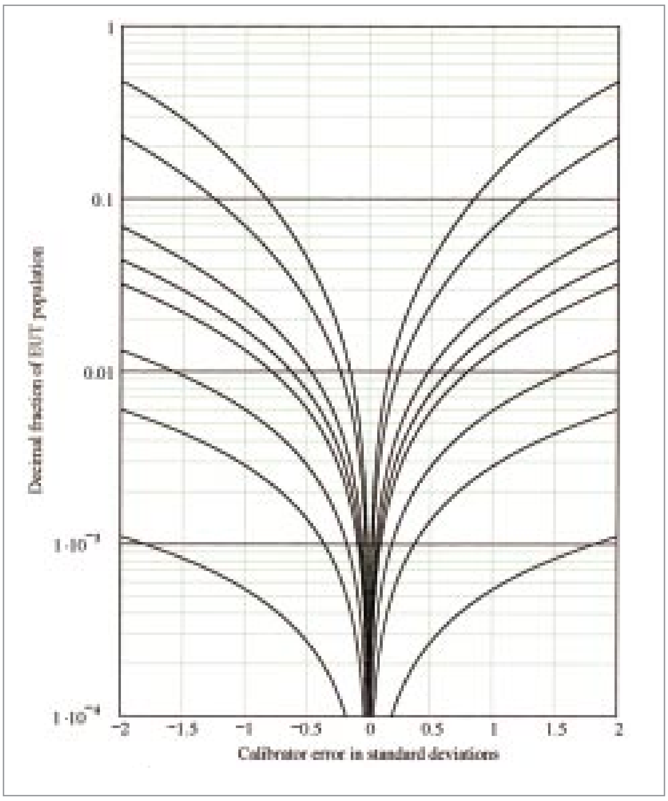



Fig. 4 Distribution of incorrect fail test decisions as a function of calibrator error for 2$\sigma$ specifications

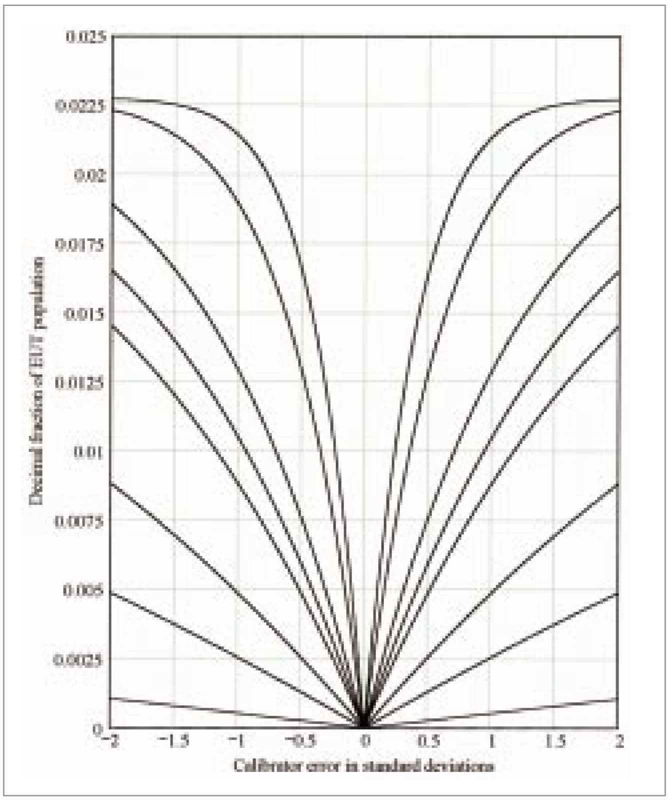



Fig. 5 Distribution of incorrect pass test decisions as a function of calibrator error for 2$\sigma$ specifications

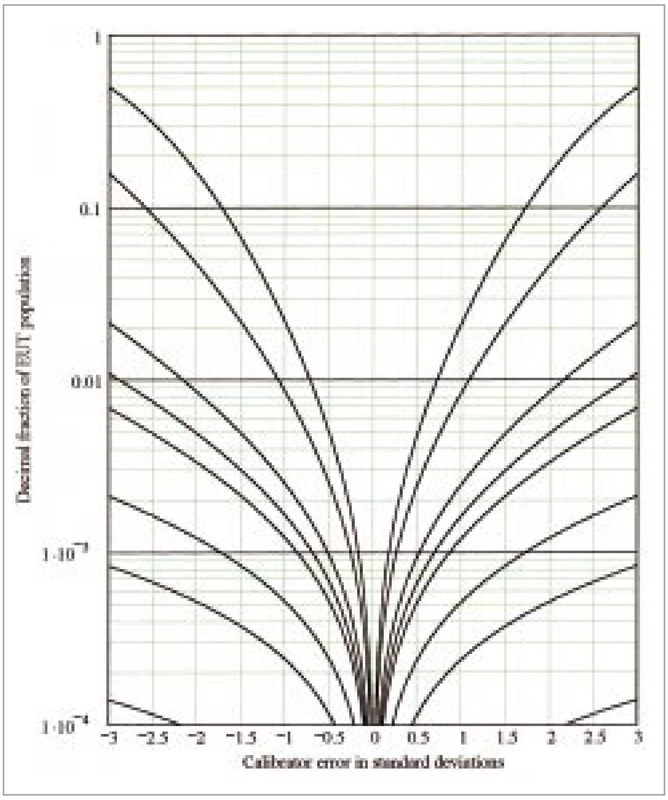



Fig. 6 Distribution of incorrect fail test decisions as a function of calibrator error for 3$\sigma$ specifications

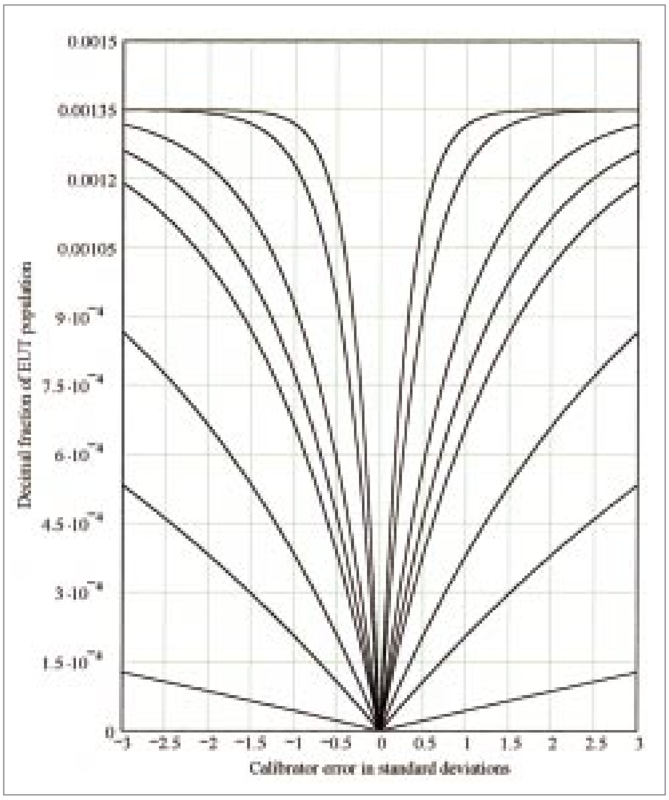



Fig. 7 Distribution of incorrect pass test decisions as a function of calibrator error for 3$\sigma$ specifications

Graphs 4 and 5 refer to 2σ specifications and graphs 6 and 7 refer to 3σ specifications for both the EUT and calibrator populations and normal distribution in the EUT population. The curves given in Figs. 4–7 refer to the following TUR values (curves from top to bottom): 1:1, 1.5:1, 3:1, 4:1, 5:1, 10:1, 20:1, 100:1. The incorrect fail unit fraction and incorrect pass unit fraction of the EUT population can be obtained from relevant values of the cumulative distribution function. It can be seen from the data in the Figures that the percentages of incorrect test decisions, and thus the percentage of correct test decisions, depend strongly on the TUR value and the confidence interval. The percentage of correct test decisions increases when the TUR or confidence interval increases.

But increasing the TUR requires the use of measuring (calibrating) instruments of higher accuracy, which can be more costly. An increase in the confidence interval increases the uncertainty of measurement. As long as the minimum TUR is met or exceeded, the uncertainties of the measurement standard when assigning an uncertainty to the calibration can be ignored.

The results of analysis indicate that 2σ confidence interval specifications require a much larger TUR value than 3σ confidence interval specifications in order to ensure the same percentage of correct test decisions. For example, assuming the 3:1 TUR, the percentage of incorrect fail test decisions is circa 6.85 % (see Fig. 4) and the percentage of incorrect pass test decisions is circa 1.89 % (see Fig. 5) for 2σ specifications when the calibrator output is just within specifications at the ± 2σ limit. For the same TUR, the percentage of incorrect fail test decisions is circa 2.14 % (see Fig. 6) and the percentage of incorrect pass test decisions is circa 0.13 % (see Fig. 7) for 3σ specifications when the calibrator output is just within specifications at the ± 3σ limit.

It is necessary to increase the TUR more than two times, i.e. to more than 6:1 for 2σ specifications if the percentage of incorrect fail test decisions is not to exceed 2.14 % too. The percentage of incorrect pass test decisions circa 0.13 % for 2σ specifications is at circa 85:1 TUR. The last condition requires using very accurate measurement standards to perform the measurement.

In some cases it is possible to find measured instruments with the uncertainty being *de facto* nearly the same as the uncertainty of the calibrating instrument used to calibrate them, i.e. the TUR is about 1:1. In the case of 2σ specifications, taking into consideration the data from Figs. 4 and 5 for 1:1 TUR, one can say that about 50 % of test decisions would be incorrect, i.e. about 47.7 % of the good EUT units would be rejected (Fig. 4) and about 2.27 % of the bad EUT units would be accepted (Fig. 5), when the calibrator output is just within specifications at the ± 2σ specification limit. Similarly, in case of 3σ specifications, the percentage of incorrect test decisions would be about 50 % too, i.e. about 49.8 % of the good EUT units would be rejected (Fig. 6) and about 0.14 % of the bad EUT units would be accepted (Fig. 7), when the calibrator output is just within specifications at the ± 3σ specification limit. As one assumes normal error distribution in the calibrator population, about 2.28 % of that population for 2σ specifications and about 0.14 % for 3σ specifications will fall under this condition.

There are some practical activities in science and technology fields where TUR values as large as 100:1 are required. Such TUR values enable a high reliability of test results and high confidence in the measurement process to be obtained. In such cases the percentage of incorrect test decisions would be as low as about 0.22 % when the calibrator error is just within specifications at the ± 2σ specification limit (see Figs. 4 and 5) for 2σ specifications and incorrect test decisions as low as about 0.027 % when the calibrator error is just within specifications at the ± 3σ specification limit (see Figs. 6 and 7) for 3σ specifications.

## 5 Two examples of a lack of specific information on the TUR

A lack of adequate or complete specific information on the TUR can be noticed even in some official documents and measurement procedures. In effect, in such cases inexperienced persons can have some difficulties in making proper measurements. For illustration, two examples concerning measurement uncertainty requirements of standards are discussed below.

### ISO 10012-1 standard [7]

The requirements on the TUR arise from clause 4.3 of this standard, which reads: "The error attributable to calibration should be as small as possible. In most cases of measurement, it should be no more than one third and preferably one tenth of the permissible error of the confirmed equipment when in use". If normal error distribution is assumed for both the EUT and the calibrating instrument then the TUR is 3:1 for the lower permissible limit of error ratio, according to the above-mentioned requirements of the standard.

Thus, even for 3σ specifications (see Figs. 6 and 7), there will be about 2.28 % of incorrect test decisions when the calibrating instrument error is just within specifications at the ± 3σ specification limit, and as much as about 8.7 % of incorrect test decisions for 2σ specifications (see Figs. 4 and 5) when the calibrating instrument error is just within specifications at the ± 2σ specification limit.

### IEC 60373 [8] and IEC 60645-1 [9] standards

The requirements on measurement uncertainty for the mechanical coupler arise from clause 5.1 of IEC 60373, which reads: "The calibration uncertainty shall not exceed 1.0 dB for frequencies up to and including 2 kHz nor shall it exceed 2 dB for frequencies up to and including 8 kHz". The mechanical coupler is a piezo-electric transducer, which is used in calibrating the stimulus level of the audiometer bone conduction. The mpe for the stimulus level of the audiometer is ± 3 dB for frequencies up to and including 4 kHz [9].

Assuming normal error distribution for the stimulus level for both the audiometer and mechanical coupler one has a 1.5:1 TUR value at 3 kHz. At this frequency, taking into consideration results of the analysis given above (see Figs. 4–7) one can draw the following conclusions. If the mechanical coupler used for calibration of audiometers and the audiometers are calibrated according to these standards, there will be about 15.9 % of incorrect audiometer test decisions for 3σ specifications, i.e. 15.9 % of incorrect rejections or incorrect acceptances of audiometers, when the error of the mechanical coupler is just within specifications at the ± 3σ specification limit and as much as about 25.2 % of incorrect audiometer test decisions for 2σ specifications when the error of the mechanical coupler is just within specifications at the ± 2σ specification limit.

## 6 Conclusions

Results of the study indicate the way in which the TUR and confidence interval affect the incorrect test decisions and thus the correct test decisions, reliability of test results and confidence in the measurement process.

Larger TUR values and confidence intervals signify lower percentages of incorrect test decisions, higher reliability of test results and higher confidence in the measurement process.

But larger TUR values require the calibrating instrument to be of higher accuracy, which usually implies a higher cost. A larger confidence interval signifies a higher uncertainty of measurement.

As long as the minimum TUR is met or exceeded, at an assumed value of confidence interval, the uncertainties of the measurement standard (or, generally, of the measuring instrument) when assigning an uncertainty to the calibration or measurement can be ignored. ■

## 7 References


[1] Finkelstein, L., Grattan, K.T.V., Eds., *Concise Encyclopedia of Measurement & Instrumentation*, Oxford, New York, Seoul, Tokyo, Pergamon Press, 1994

[2] Sydenham, P.H., Ed., *Handbook of Measurement Science*, Vol. 1, Chichester, New York, Brisbane, Toronto, Singapore, John Wiley & Sons, 1982

[3] *International Vocabulary of Basic and General Terms in Metrology (VIM)*, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML - ISO, Geneva, 1993

[4] *Guide to the Expression of Uncertainty in Measurement (GUM)*, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML - ISO, Geneva, 1995 (Corrected & reprinted edition)

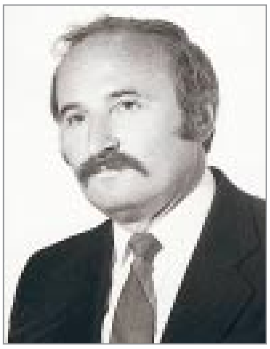
[5] Turzeniecka D., *Study of Results of Comparison of Selected Uncertainties*, Metrology and Measuring Systems, Vol. V, 1-2, PWN, Warsaw, 1998

[6] *Calibration: Philosophy in Practice*, 2nd Ed., Fluke Corporation, 1994

[7] ISO 10012-1, *Quality assurance requirements for measuring equipment-Part 1: Metrological confirmation system for measuring equipment*, ISO, Geneva, 1993

[8] IEC 60373, *Mechanical coupler for measurements on bone vibrators*, IEC, Geneva, 1990

[9] IEC 60645-1, *Pure-tone audiometers*, IEC, Geneva, 1992


TADEUSZ SKWIRCZYŃSKI

Independent Consultant,
Warsaw, Poland

The author welcomes comments from readers and may be contacted by e-mail: tadskwi@wp.pl