

# Proficiency testing for the calibration of masses

Lise HEGRON<sup>1,\*</sup>, Boris GEYNET<sup>2</sup>

<sup>1,2</sup> CT2M, Route de Lançon, Centre des Creusets, 13250 Saint-Chamas, France

**Abstract.** The CT2M organized in 2018 a european inter-laboratory comparison (ILC) for the calibration of masses. The proficiency testing was particularly intended for the calibration laboratories (accredited or not) but also the testing laboratories carrying out their own calibrations and / or controls of their masses. This circuit took place between April 2018 to October 2018 in five European countries: England, France, Germany, Portugal and Switzerland. The results were processed according to the statistical principle of ISO 13528 [1] and in compliance with the requirements of ISO 17043 [2]. This article presents the organization of this inter-laboratory comparison and the results. The performances of the participants are evaluated and an interpretation of the results is proposed in order to highlight the predominant influence parameters on the mass calibration results (nominal values: 200mg, 2g, 20g, 200g and 20kg).

## 1 Introduction

Accreditation body require accredited laboratories to regularly participate in inter-laboratory comparisons (ILC) to prove their ability to perform tests or calibrations. Since 2014 CT2M organized inter-laboratory comparison in various fields to meet that need. In 2018, an inter-laboratory comparison was organized for the calibration of masses. 15 laboratories took part in this round which took place from April 2018 to October 2018. The participants were calibration laboratories (accredited or not) as well as testing laboratories carrying out internal calibrations and / or controls of their masses.

This proficiency testing scheme was organized in accordance with the requirements of ISO 17043 and the results processing and participant performance study were conducted according to the principles of ISO 13258. The results obtained are presented in this paper.

## 2 Organisation of the proficiency texting round

### 2.1 Proficiency testing items

The calibration items consisted in 5 masses:

- ✓ 4 masses (OIML class E2) of nominal values: 200mg, 2g, 20g and 200g
- ✓ 1 mass (OIML class F1) of nominal value 20kg.

---

\* Corresponding author: [lhegron@ct2m.fr](mailto:lhegron@ct2m.fr)

To ensure the stability of the masses during the ILC round, an accredited reference laboratory, which was not part of the participants, calibrated the masses at the beginning and at the end of the round.

## 2.2 Procedure of the proficiency testing

The preferred calibration method was the reference method of comparing the mass to be calibrated with a standard weight of equivalent nominal value, using a comparator. Each laboratory used its procedure and was free to choose the number of repetitions and the number of calibration cycles (Table 1).

**Table 1.** Number of repetitions and calibration cycles

Lab	Number of repetitions and cycles (A: standard mass ; B: calibrated mass)				
	200mg	2g	20g	200g	20kg
1	10 ABAB	10 ABAB	10 ABAB		
2	5 ABBA		5 ABBA	5 ABBA	5 ABBA
3	1 ABBA	1 ABBA	1 ABBA	1 ABBA	2 ABBA
4	3 ABBA	3 ABBA	3 ABBA	3 ABBA	3 ABBA
5		5 AABB	5 AABB	5 AABB	
6		2 ABBA	2 ABBA	2 ABBA	2 ABBA
7	2 ABBA	2 ABBA	2 ABBA	2 ABBA	2 ABBA
8	2 ABBA	2 ABBA	2 ABBA	2 ABBA	2 ABBA
9	3 ABBA	3 ABBA	3 ABBA	3 ABBA	3 ABBA
10	3 ABBA	3 ABBA	3 ABBA	3 ABBA	3 ABBA
11	5 ABA	5 ABA	5 ABA	5 ABA	3 ABA
12	2 ABBA	2 ABBA	2 ABBA	2 ABBA	2 ABBA
13	2 ABABA	2 ABABA	2 ABABA	2 ABABA	2 ABA
14			10 ABBA	10 ABBA	
15	4 ABBA	4 ABBA	4 ABBA	4 ABBA	

The participants used their own means of calibration, namely:

- ✓ Standard masses of nominal values equivalent to the masses to be calibrated,
- ✓ A scale or a comparator.

The characteristics of the standard masses, of the scale / comparator and of environmental conditions had been indicated by the participants.

The participants had to determine the conventional mass of each calibrated mass. In addition, the following information could be mentioned :

- ✓ The expanded calibration uncertainty ( $k=2$ ),
- ✓ Influence factors considering for the uncertainty calculation,
- ✓ Environmental conditions during calibration.

### 3 Method for data analysis

#### 3.1 Hypothesis for data analysis

The purpose of the proficiency testing is to conduct an assessment of laboratory performance by comparing their results with each other and against a reference value. Assumptions were considered for the data analysis to make consistent conclusions on the results.

In accordance with the reference standards [1] [3] [4], the assumption of a normal distribution of the data series has been made. This hypothesis has been verified for all results using the following two methods:

- ✓ Graphical Method (normal probability plot)
- ✓ Statistical method (Shapiro-Wilk)

The Grubbs test was performed on all the conventional masses determined by each of the participant. The purpose of this test is to identify a laboratory with an incoherent result compared to other participants. The test involves calculating the Grubbs parameter (G) and comparing it to the critical Grubbs values in the Table 5 of ISO 5725-2 [3].

Only one outlier was highlighted by the Grubbs test. It was the conventional mass of the weight of 2g obtained by the laboratory No.13 whose value was abnormally low (Fig. 3a).

#### 3.2 Assigned values and standard deviations

##### 3.2.1 Reference value and uncertainty

The reference values of this proficiency testing were obtained by a reference laboratory that is accredited according to ISO/IEC 17025. It calibrated the masses at the beginning and at the end of the ILC round. The reference value  $x_{\text{ref}}$  is the average of both conventional masses. Its associated uncertainty  $U_{\text{ref}}$  takes into account the uncertainty of the reference laboratory as well as the possible drift of the masses between the beginning and the end of the round.

**Table 2.** Reference values and uncertainties

Mass	Reference values	
	$x_{\text{ref}}$ (g)	$U_{\text{ref}}$ (g) k=2
200mg	0,2000038	0,0000062
2g	1,999987	0,000013
20g	19,999937	0,000026
200g	199,99977	0,00011
20kg	20000,016	0,031

The table 2 lists the reference values and their associated uncertainties.

##### 3.2.2 Robust mean and robust standard deviation

The robust mean  $x^*$  and the robust standard deviation  $S^*$  are determined using the algorithm A defined in ISO 13528. The robust average and standard deviation are used to

evaluate the Z-score of each participant. The Table 3 shows the parameters obtained for all the tests, *ie.*:

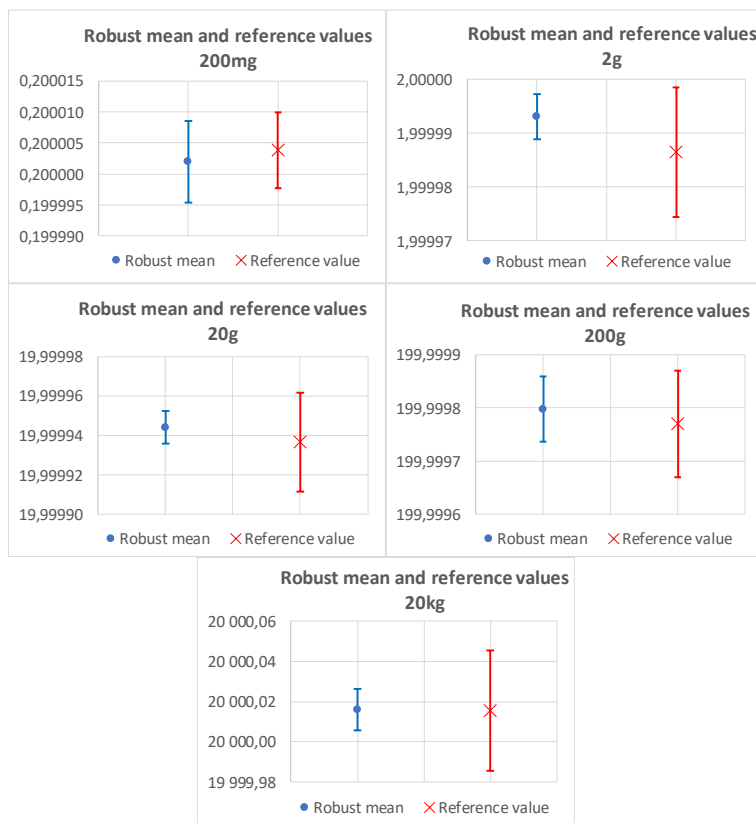
- $\bar{x}^*$ : robust average
- $S^*$ : robust standard deviation
- $U_{\bar{x}^*}$ : uncertainty on the robust average

**Table 3.** Robust mean and standard deviation

Mass	Number of lab.	Robust mean and standard deviation		
		$\bar{x}^*$ (g)	$S^*$ (g)	$U_{\bar{x}^*}(g) k=2$
200mg	12	0,2000020	0,0000092	0,0000067
2g	13	1,9999930	0,0000061	0,0000043
20g	15	19,9999440	0,0000130	0,0000085
200g	14	199,999798	0,000092	0,000063
20kg	11	20000,016	0,014	0,011

### 3.2.3 Comparison between the robust averages and the reference values

When organizing an inter-laboratory comparison, it is important to ensure that the robust averages of the participants results are not significantly different from the reference values. The graphs below show a good correspondence between the robust means and the reference values for each of the calibrated masses.



**Fig. 1.** Comparison between the robust means and the reference values

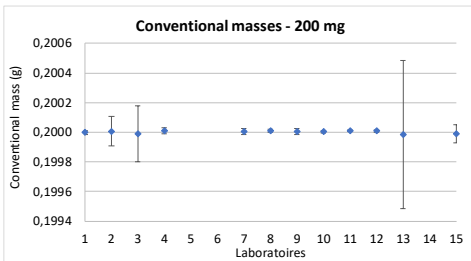
The number  $En$  highlights the difference between the robust average with its uncertainty and the reference value with its uncertainty (Table 4). The number  $En$  between both the results are in the range  $[-1; 1]$  for all the calibrated masses. It is therefore possible to conclude that there is no significant difference between the reference values (obtained by the reference laboratory) and the robust averages (participants results).

**Table 4.** Number  $En$

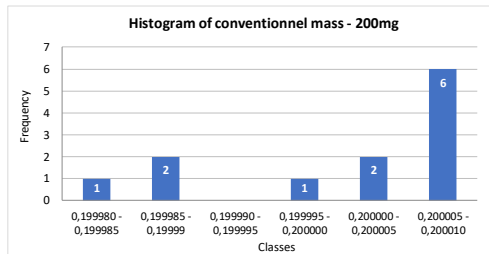
Mass	Number of lab.	$En$
200mg	12	-0,20
2g	13	0,51
20g	15	0,28
200g	14	0,24
20kg	11	0,01

### 4 Participant results

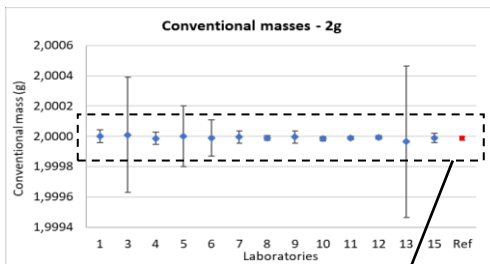
The results for the 5 masses are presented in the following graphs. They show the conventional masses and their uncertainties associated with  $k=2$  (error bars). For each calibration point, a histogram also shows the frequency of the results (number of laboratories) according to the class of values.



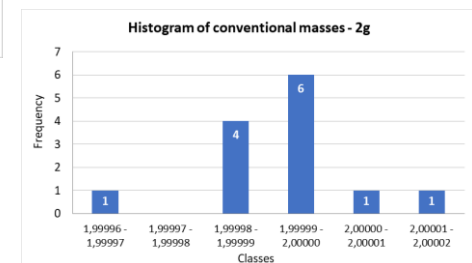
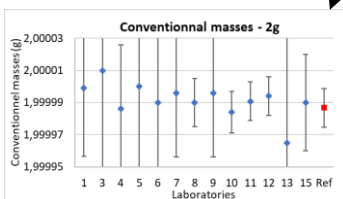
**Fig. 2a.** Conventional masses and their uncertainties for 200mg



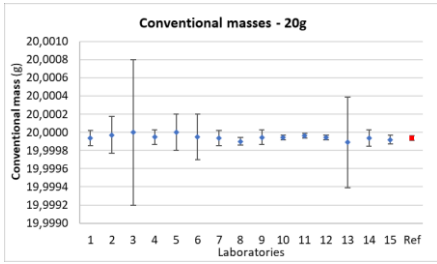
**Fig. 2b.** Histogram of conventional mass for 200mg



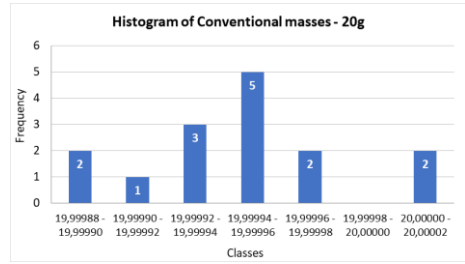
**Fig. 3a.** Conventional masses and their uncertainties for 2g



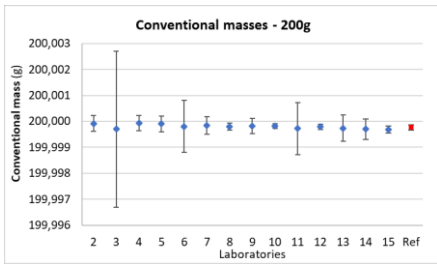
**Fig. 3b.** Histogram of conventional masses for 2g



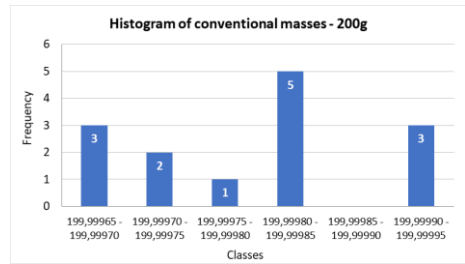
**Fig. 4a.** Conventional masses and their uncertainties for 20g



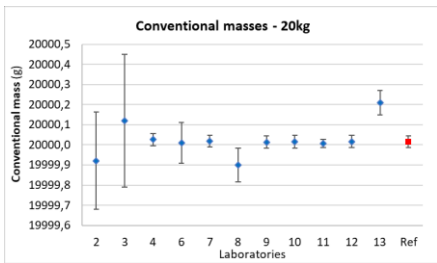
**Fig. 4b.** Histogram of conventional masses for 20g



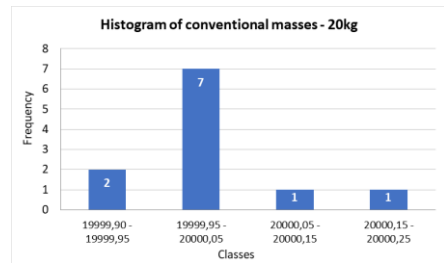
**Fig. 5a.** Conventional masses and their uncertainties for 200g



**Fig. 5b.** Histogram of conventional masses for 200g



**Fig. 6a.** Conventional masses and their uncertainties for 20kg



**Fig. 6b.** Histogram of conventional masses for 20kg

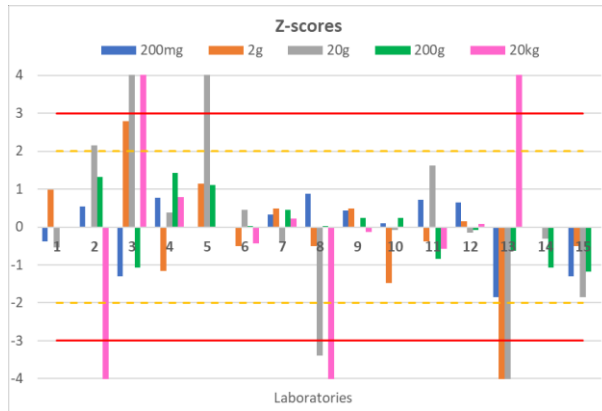
## 5 Participant performance

The laboratories performance is determined by the Z-score that is a standardized measure of the bias. This performance score is calculated using the following formula:

$$Z = \frac{(x_{lab} - x^*)}{S^*} \tag{1}$$

The reference value  $x^*$  is the robust average of the participants results, the value  $x_{lab}$  is the value obtained by the laboratory and the parameter  $S^*$  is the robust standard deviation of the participants results.

For each participant, the z-scores were calculated for each calibrated mass (Fig. 7). The Z-scores between -2 and -3 or 2 and 3 correspond to isolated result. The Z-scores less than -3 or greater than 3 correspond to discordant result.



**Fig. 7.** Z-scores of each participant for all the calibrated masses

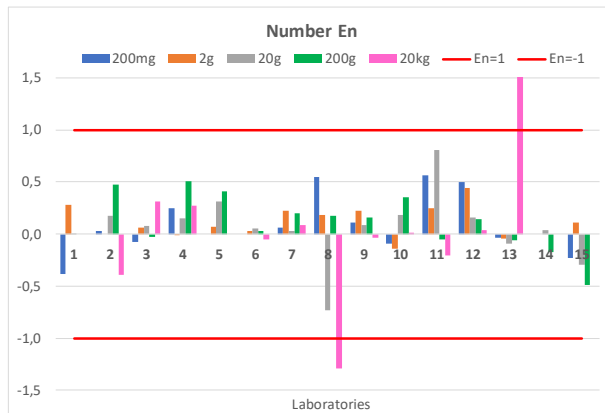
If a laboratory obtains a positive (or negative) *Z*-score for all the calibrated masses, this highlights an overall bias (systematic error) on the calibration process compared to the average of the participants. This bias is all the more so high as the values of *Z*-scores is far from 0.

An interpretation of the results in relation to the reference value was also performed using the number *En*. It is a parameter that lead to evaluate wthe difference between 2 values. The values are compared taking into account their associated expanded uncertainties. The results of the participants are therefore compared to the reference value. The number *En* is calculated using the following formula:

$$En = \frac{x_{lab} - x_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}} \tag{2}$$

- $x_{ref}$  is the reference value (reference laboratory),
- $x_{lab}$  is the laboratory value,
- $U_{ref}$  is the uncertainty ( $k=2$ ) on the reference value,
- $U_{lab}$  is the uncertainty ( $k=2$ ) on the laboratory value.

The Figure 8 shows the results of all the numbers *En* for the different calibrated masses. If a laboratory obtains a positive (or negative) number *En* on all the calibrated masses, this highlights an overall bias (systematic error) on the calibration process compared to the reference value. This bias is all the more so important as the number *En* is far from 0.



**Fig. 8.** Numbers  $En$  for all the masses

The interpretation of the number  $En$  must be done with caution. Indeed some laboratories have significant uncertainties, which leads to a correct standardized deviation (less than 1 or more than -1) despite significant bias compared to the reference value (laboratory No.3 for example).

## 6 Conclusions

This inter-laboratory comparison bringing together a sufficient number of laboratories made it possible to highlight several conclusions.

The results of the participants were exploited and performance criteria were provided to participants so that they could either validate their calibration method as accreditation bodies, or improve it by triggering actions to correct a possible bias.

It should be noted that several accredited laboratories participated in this inter-laboratory comparison have results that are significantly different from the assigned values. Moreover, no correlation can be made with the country of the laboratories.

## References

1. ISO 13528 October 2015: Statistical methods for use in proficiency testing by interlaboratory comparisons
2. EN ISO/CEI 17043 April 2010: Conformity assessment - General requirements for proficiency testing
3. ISO 5725-2 December 1994 : Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method