CIPM MRA
Comparison reports

EURAMET.L-S2.2.n01

Gauge block calibration by mechanical comparison

SUPPLEMENTARY COMPARISON

© 2026, J. S. Eensalu Hall *et al*
This report is published by the BIPM.
Original content from this Report may be used under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) Licence.

Any further distribution of this Report must be cited as:
J. S. Eensalu *et al* 2026 *CIPM MRA Comparison reports* **04008**
https://doi.org/10.59161/HBFM4577

The CIPM MRA Comparison reports are made available under the Creative Commons Attribution International licence:
Attribution 4.0 International (CC BY 4.0)

By using this Report, you accept to be bound by the terms of this licence
(https://creativecommons.org/licenses/by/4.0/).

Distribution – you may distribute the Report according to the stipulations below.

Attribution – you must cite the Report.

Adaptations – you must cite the original Report, identify changes to the original and add the following text: This is an adaptation of an original Report by the Author(s). The opinions expressed and arguments employed in this adaptation should not be reported as representing the views of the Authors.

Translations – you must cite the original Report, identify changes to the original and add the following text: In the event of any discrepancy between the original work and the translation, only the text of the original Report should be considered valid.

Third-party material – the licence does not apply to third-party material in the Report. If using such material, you are responsible for obtaining permission from the third-party and of any claims of infringement.

Supplementary Comparison EURAMET.L-S2.2.n01

Calibration of gauge blocks by mechanical comparison

(EURAMET project 1581)

Final Report

AS Metrosert

Central Office of Metrology

Jako Siim Eensalu, Metrosert

Coordinator: Allar Pärn, Metrosert

Tallinn, June 2026

Contents

1 Document control 3
2 Introduction 5
3 Organization 6
3.1. Participants 6
3.2. Schedule 7
3.3. Standards 7
4 Measurement instructions and reporting of results 8
5 Stability and condition of the gauge blocks 10
6 Laboratory measurements 12
7 Measurement results 12
7.1. Results and standard uncertainties as reported by participants 12
7.2. Measurement uncertainties 19
8 Analysis 21
8.1. Calculation of the SCRV, Degrees of Equivalence, and Chi-squared 21
8.2. Participant responses 34
8.3. Analysis of revised DPM results 42
8.4. Discussion of results 43
9 Conclusions 44
Appendix A – Equipment and measuring processes of the participants 45
10 References 51

1 Document control

Version	
Draft A.1	*Issued on May 29, 2025.*
Draft A.2	*Issued on January 28, 2026.* Revised Tables Table 13, Table 14, Table 15, Table 17. Added missing decimals to Metrosert results in Table 4 through Table 9. Added “Participant responses” chapter. Corrected absolute expanded uncertainty CMC entry equation for BMM from quadratic to linear in Table 10. Added the pilot’s measured and time-averaged deviation from nominal central length values to Table 11 and variation values to Table 12 per A. Lewis’ request. DPM reported on Sept. 17, 2025 that due to a technical error in Excel processing, i.e. $L_c = L_n + R_2 - R_1$ instead of $L_c = L_{ref} + R_2 - R_1$, all calibration results from DPM were stated to not be traceable to the EIM reference values. Hence, per the recommendation from the TC-L chair citing T. Coveney, DPM was excluded from the RV calculation in deviation from nominal central length, and the corrected data was assessed separately. Added Table 20, Figure 12.
Draft A.3	*Issued on March 12, 2026* The standard uncertainty values reported by BMM were previously mistakenly divided by 2 (to k = 0,5) in the report, as reminded by T. Božović. Thus, SCRV of deviation from nominal central length and variation was recalculated for all gauge blocks with the corrected BMM uncertainty values. All graphs and tables containing results by BMM were revised. Added Table 19 to summarise the statistical analysis. Revised Table 20, Figure 12 to reflect the updated SCRV. Corrections to minor style and omission mistakes, as noted by A. Lewis. Added (k = 1) to chapter 8.1 text to emphasize standard uncertainties were used as input in all analysis calculations. Revised the analysis description to include chi-squared analysis for participants who failed the E_n evaluation as per recommendation by A.

	Lewis. NMI was assigned $u(k = 1)$ of Q[24,3;0,42 L] based on CMC entry data for chi-squared analysis, as length dependent uncertainty components were not provided by NMI in the supplied uncertainty budget. Birge ratio analysis was corrected to use the dynamic 95 % confidence upper limit for the Birge ratio cutoff instead of 1,0.
Draft B.1	*Issued on April 8, 2026* No changes since A.3.
Draft B.2.	*Issued on June 11, 2026* Corrected spelling mistakes on pages 23, 25, and 43.

2 Introduction

The metrological equivalence of national measurement standards and of calibration certificates issued by national metrology institutes is established by a set of key and supplementary comparisons chosen and organized by the Consultative Committees of the CIPM or by the regional metrology organizations in collaboration with the Consultative Committees.

In 2022, at the annual EURAMET TC-L meeting it was proposed and agreed upon to have a supplementary comparison on gauge block calibration by mechanical comparison. The purpose of this supplementary comparison is to support new and existing CMC claims for this method.

For this supplementary comparison numbered EURAMET.L-S2.2.n01 Metrosert ltd acted as the pilot laboratory and NPL acted as the co-pilot. Metrosert was responsible for providing the travelling standards, monitoring the stability of travelling standards, gathering the measurement data and evaluation and reporting of the measurement results. NPL was responsible for performing additional interferometric measurements to provide a traceability link to a gauge block key comparison. The results of this international comparison contribute to the Mutual Recognition Arrangement (MRA) between the national metrology institutes of the Meter Convention [1].

3 Organization

3.1. Participants

All members of EURAMET TC-L were invited to participate. A total of 11 institutions decided to take part. The list of participants is given in Table 1.

Table 1. List of participant laboratories and their contacts.

COUNTRY	CONTACT	ADDRESS	PHONE, FAX, E-MAIL
ALBANIA	Defrim Bulku	DPM Autostrada Tirane-Durres Km 8 Tirana	+355 4 2233 174 defrim.bulku@dpm.gov.al
AUSTRIA	Michael Matus	BEV Arltgasse 35 1160 Wien	+43 1 21 110 6540 michael.matus@bev.gv.at
BULGARIA	Denita Tamakyarska	BIM 52B, G.M. Dimitrov blvd. 1040 Sofia	+359 2 970 27 19 d.tamakjarska@bim.government.bg
ESTONIA	Jako Siim Eensalu	METROSERT Teaduspargi 8, 12618 Tallinn	+372 5340 7617 jako-siim.eensalu@metrosert.ee
HUNGARY	Gábor Szikszai	BFKH Németvölgyi út 37-39 1534 Budapest	+36 1 4585854 szikszai.gabor@bfkh.gov.hu
IRELAND	Rory Hanrahan	NSAI NML Claremont Avenue, Glasnevin Dublin 9	+351 1 8082611 rory.hanrahan@nsai.ie
LATVIA	Larisa Svedova	LATMB Kr. Valdemara street 157 1013 Riga	+371 67 51 77 27 larisa.svedova@latmb.lv
LITHUANIA	Lilija Chaleckiene	FTMC Sauletekio av. 3, LT-10257 Vilnius	+370 61216739 +370 5 262 7123 lilija.chaleckiene@ftmc.lt
MOLDOVA	Alexandru Braguta	NMI (MD) Eugen Coca street 28 2064 Chisinau	+373 79 004 982 alexandru.braguta@inm.gov.md
MONTENEGRO	Tomo Božović	BMM Arsenija Boljevića bb 81000 Podgorica	+382 20 601 360 tomo.bozovic@metrologija.gov.me
SLOVENIA	Bojan Acko	MIRS/UM-FS/LTM Smetanova ulica 17 2000 Maribor	+386 2 220 7581 bojan.acko@uni-mb.si
UNITED KINGDOM	Andrew Lewis	NPL Hampton Road TW11 0LW Teddington, Middlesex	+44 208 943 6074 andrew.lewis@npl.co.uk
COORDINATOR	Allar Pärn	METROSERT Teaduspargi 8, 12618 Tallinn	+372 5388 9335 allar.parn@metrosert.ee tc-length.contact@metrosert.ee

3.2. Schedule

The comparison was initially organized in two loops with each laboratory allowed approximately one month to make their measurements and to prepare for transportation to the next participant. The schedule was tailored to satisfy the laboratories for scheduling the measurements. Any changes to the schedule, after the start of the circulation, were discussed and agreed among the participants and the TC-L chairman. A third loop was added to allow BMM and NSAI NML to repeat their measurements after problems with experimental setups and facility HVAC had been solved. The co-pilot laboratory, NPL, provided the traceability link of the artefacts to a gauge block key comparison by inferferometric measurements. Most delays in the schedule are attributed to improper packaging and expiration of the ATA Carnet, and the addition of two unplanned measurement slots.

Table 2. Schedule of the comparison.

Institute	Country	Original schedule	Date of measurement	Results received
METROSERT	EE	July 2023	July 2023	
BEV	AT	August 2023	August 2023	September 2023
BIM	BG	October 2023	October 2023	December 2023
FTMC	LT	November 2023	November 2023	January 2024
METROSERT	EE	January 2024	January 2024	
MIRS/UM-FS/LTM	SI	January 2024	February 2024	February 2024
NMI (MD)	MD	February 2024	March 2024	May 2024
DPM	AL	March 2024	April 2024	April 2024
LATMB	LT	April 2024	June 2024	July 2024
NPL	GB	May 2024	July 2024	July 2024
BFKH	HU	June 2024	July 2024	October 2024
METROSERT	EE	July 2024	September 2024	
NSAI NML	IE	September 2023	October 2024	April 2025
BMM	ME	December 2023	January 2025	April 2025
METROSERT	EE	-	April 2025	

3.3. Standards

The circulated artefacts were 6 steel gauge blocks of grade K of rectangular cross section, nominally according to the standard ISO 3650. As the gauge blocks were brand new, the artefact instability due to unknown long-term drift had to be measured during the comparison.

Table 3. Steel gauge block artefacts.

Identification	Nominal length (mm)	Manufacturer
218004	0.5	Mitutoyo
217026	4	Mitutoyo
222541	10	Mitutoyo
221877	25	Mitutoyo
221469	50	Mitutoyo
221625	100	Mitutoyo

4 Measurement instructions and reporting of results

Before calibration, the measurement surfaces of the gauge blocks had to be inspected for damage. Any damage, rust or scratches had to be documented in the drawing in appendix B.

The gauge blocks had to be measured by mechanical comparison with the participant laboratory's reference gauge blocks, using the normal calibration procedure. Interferometric measurements of the artefacts were prohibited.

The "A" surface was the marked measuring face for gauge blocks with nominal length < 6 mm and the right-hand measuring face for gauge blocks with nominal length ≥ 6 mm, respectively (see Figure 1). Measurements had to be done in one orientation with face "A" surface uppermost and face "B" facing the operator.

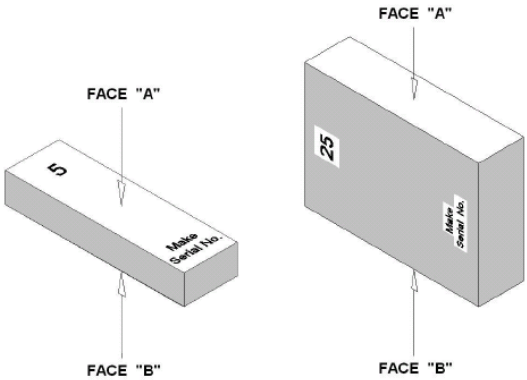

Figure 1. Gauge measurement position.

The mechanical calibration was to be performed as a five-point measurement (Figure 2) and the measurands are as stated in ISO 3650.

The following were the measurands for each gauge block: the deviation of the central length, l_c, from the nominal length, l_n, thus, $\Delta l_c = l_c - l_n$, and variation in length, v (Figure 3).

Figure 2. Gauge block measurement points.

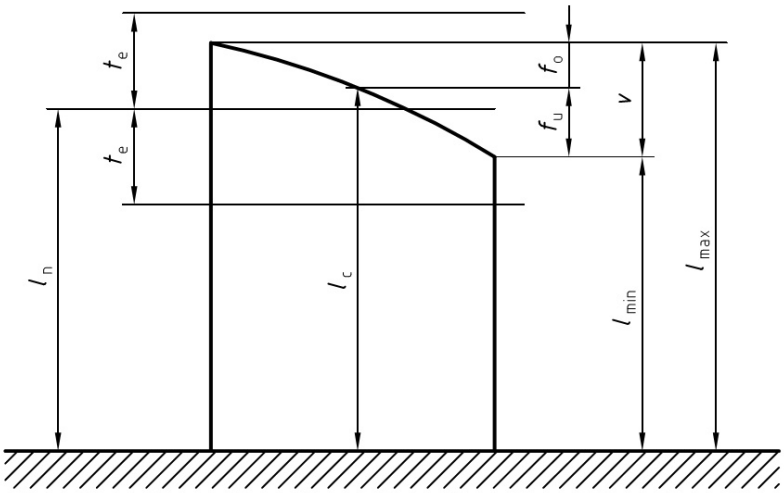

Figure 3. Nominal length (l_n), central length (l_c), variation (v) of gauge blocks.

The uncertainty of measurement was to be estimated according to the ISO *Guide to the Expression of Uncertainty in Measurement*. The participating laboratories were encouraged to use their usual model for the uncertainty calculation.

The participants were asked to report the standard uncertainty for each deviation from nominal central length $u(\Delta l_c)$ and standard uncertainty for each variation in length $u(v)$, and the uncertainty budget of the resulting combined uncertainty.

In case of relevant CMC-s the expanded measurement uncertainty for deviation from nominal central length from the nominal length $U(\Delta l_c)$ had to be expressed in the usual length-dependent form:

$$U(\Delta l_c) = Q[a, b \times l_n] = \sqrt{a^2 + (b \times l_n)^2}$$

Measurement results had to be reported for the reference conditions as per the standard ISO 3650 with uncertainties calculated according to GUM (JCGM 100:2008) [2].

5 Stability and condition of the gauge blocks

The central length of the gauge blocks was relatively stable during the 625-day long journey. As seen in Figure 4, the central length slightly increased for the 0,5 mm gauge block and decreased for the other five over time, remaining within the uncertainty of the pilot laboratory. The variation increased for all gauge blocks over time (Figure 5), indicating an increase in surface roughness due to mechanical damage during contact measurements. Thus, least squares linear regression was applied to evaluate the drift of the reference value. Accordingly, the gauge blocks are all considered stable, meaning the standard evaluation procedure was not altered other than calculating the pilot reference value as a time-weighted average.

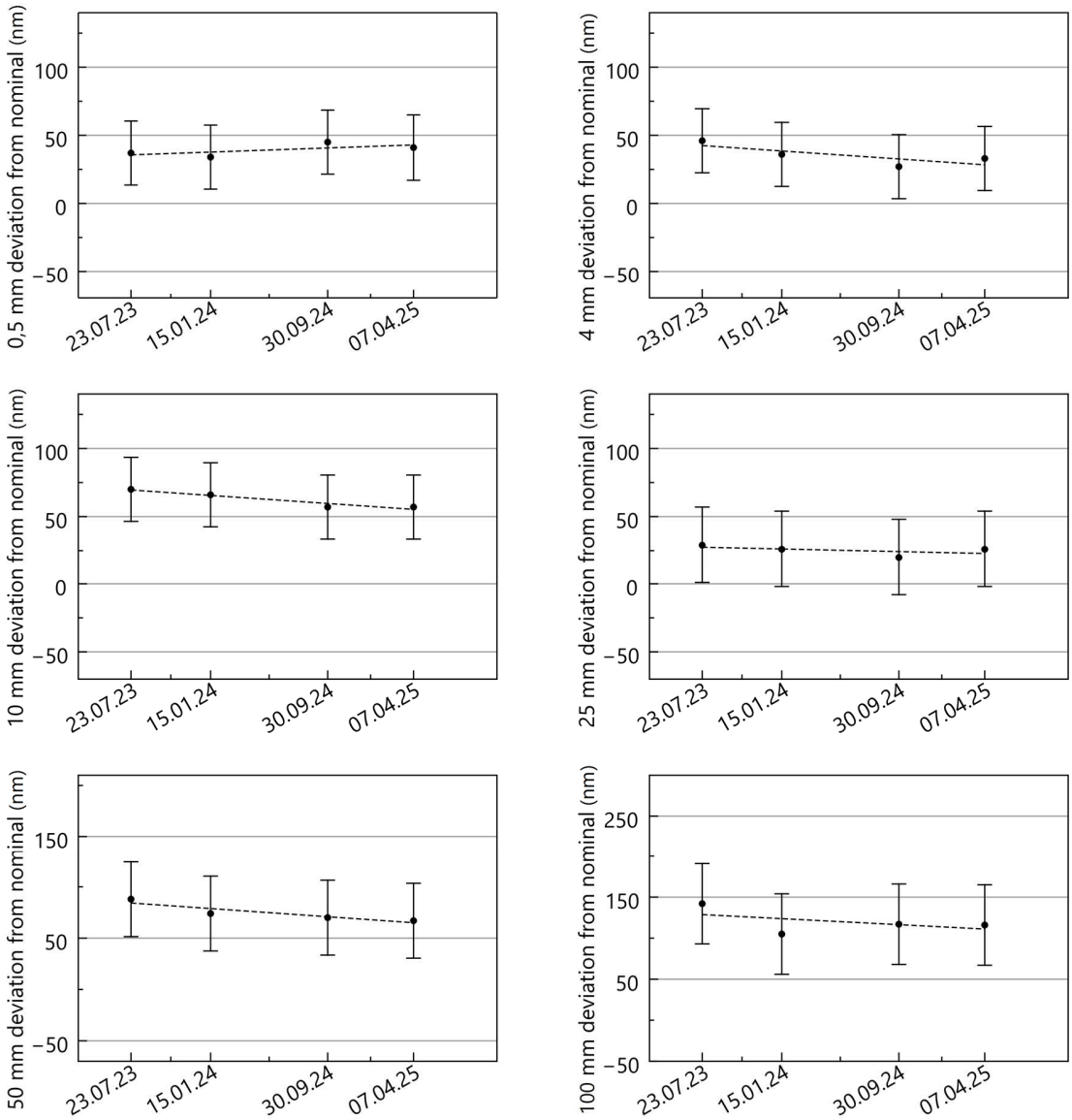

Figure 4. Drift of the deviation from nominal central length of the gauge blocks (k = 1).

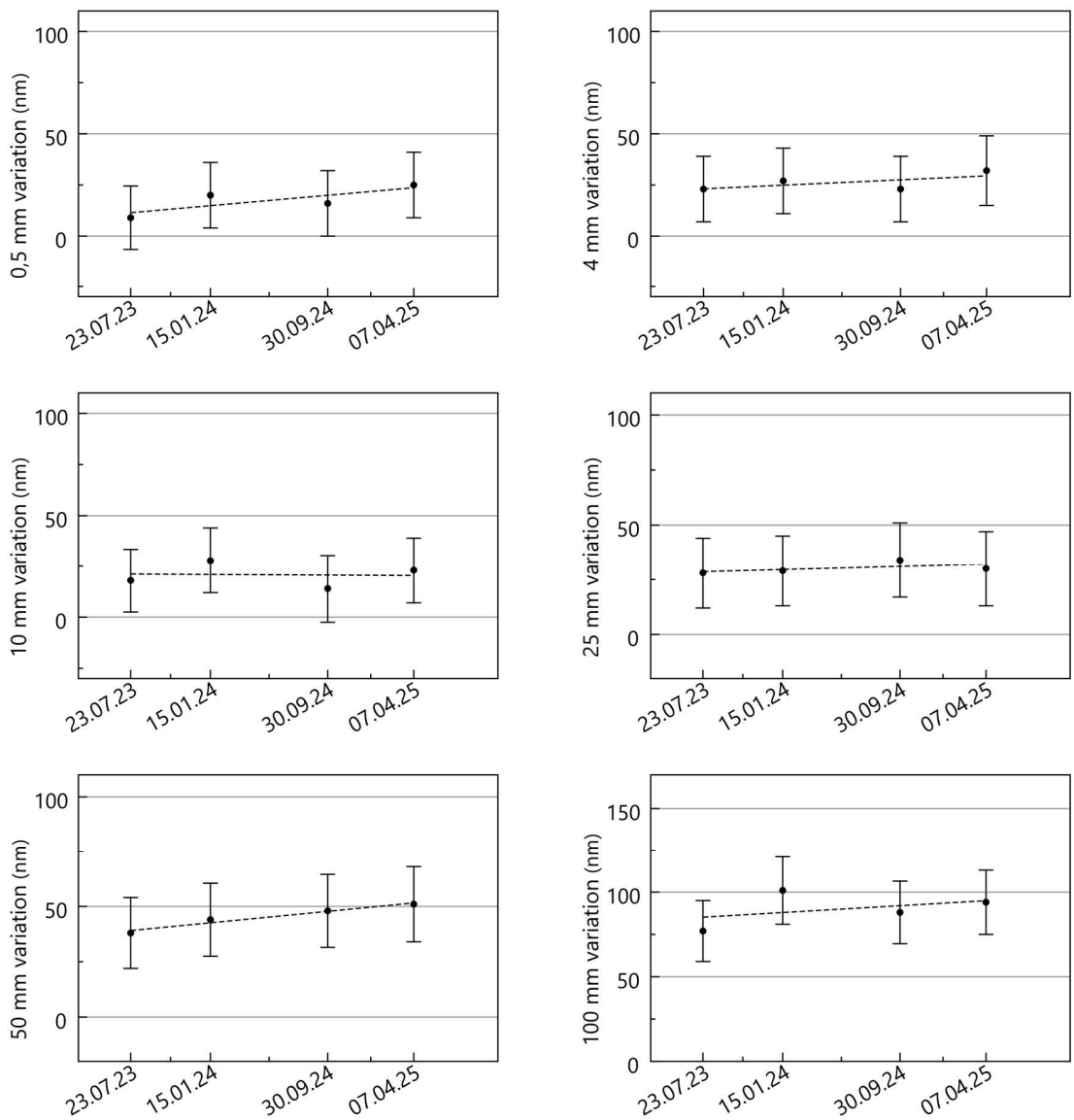

Figure 5. Variation of the gauge blocks (*k* = 1) over time.

At the beginning of the comparison, most of the gauge blocks were in pristine condition, yet some showed signs of use: the center of face A on the 4 mm gauge block was scratched, and face B of the 100 mm gauge block had four scratches running parallel to the longer edge of the face. During the comparison, all but one participant noted after visual inspection in Appendix A that there were no apparent damages, whereas BMM stated to have detected severe damages putting the measurement results at risk.

All six gauge blocks were reported to have visible scratches on faces A and B, some of which participants attributed to wringing, as some of the streaks ran parallel to the long edge of the measuring faces. Based on observations by multiple participants, evidence provided in Appendix B, and subsequent visual inspection by the pilot laboratory at the end of the

comparison, suggests the 100 mm gauge block had been slid across the comparator surface at some point during the first year of the comparison while the lower mechanical probe was touching the gauge block, leaving a straight line across the center of face B.

6 Laboratory measurements

The principal equipment used by the participants was a selection of gauge block comparators manufactured by Mahr (5 participants), TESA (4 participants), Cary (2 participants), and Feinmess Suhl (1 participant). The largest reported temperature deviation from the reference 20 °C was +0,6 °C.

7 Measurement results

7.1. Results and standard uncertainties as reported by participants

The results reported by the participants are given in tables Table 4 through Table 9 and figures Figure 6 through Figure 11 show the reported deviations from nominal length, with standard (k = 1) uncertainties.

Table 4. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **0,5 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	37	23,5	9	15,5
BEV	21	25	14	15
BIM	33	50	40	55
FTMC	45	22	38	22
METROSERT	34	235	20	16
MIRS/UM-FS/LTM	30	18	30	20
NMI	92	30	10	24
DPM	-13	27	8	24
LATMB	10	26	20	26
NPL	53	16	30	20
BFKH	22	20	49	31
METROSERT	45	23,5	16	16
NSAI NML	18	10	20	10
BMM	40	26	20	26
METROSERT	41	24	25	16

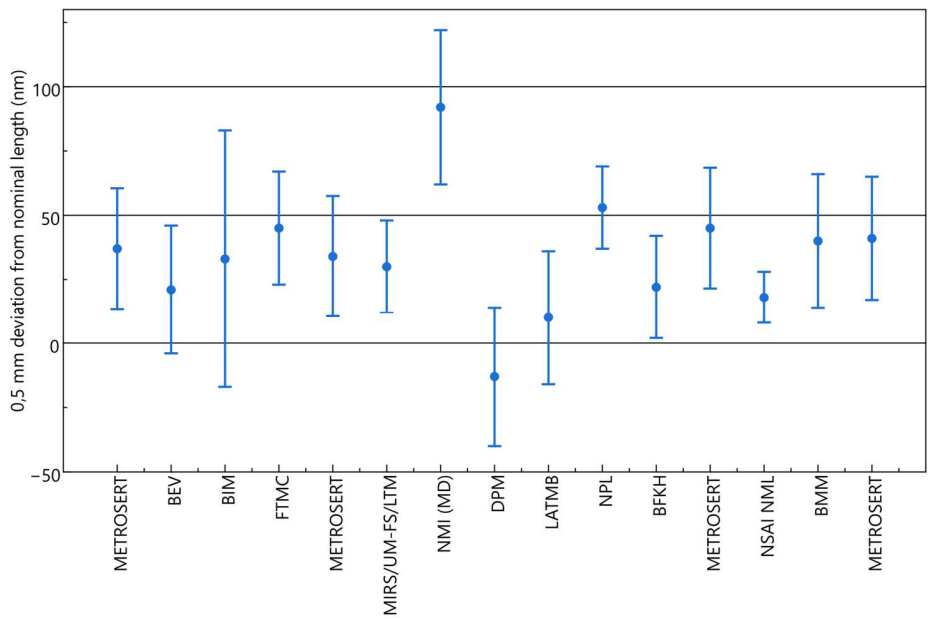

Figure 6. Deviation from nominal central length of the 0,5 mm steel gauge block (k = 1).

Table 5. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **4 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	46	23,5	23	16
BEV	13	25	32	15
BIM	-2	50	30	55
FTMC	15	22	20	22
METROSERT	36	23,5	27	16
MIRS/UM-FS/LTM	30	18	5	20
NMI	52	30	2	24
DPM	41	27	28	24
LATMB	10	26	20	26
NPL	72	17	43	20
BFKH	6	19	37	29
METROSERT	27	23,5	23	16
NSAI NML	42	10	20	10
BMM	30	26	40	26
METROSERT	33	23,5	32	17

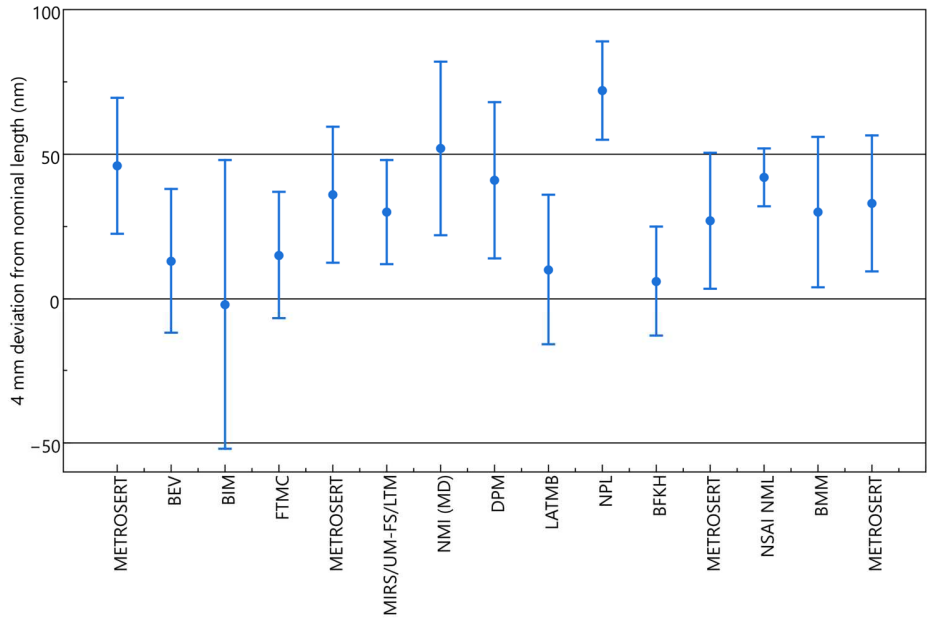

Figure 7. Deviation from nominal central length of the 4 mm steel gauge block (k = 1).

Table 6. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **10 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	70	26,5	18	15,5
BEV	30	25	37	15
BIM	-17	50	30	55
FTMC	26	22	10	22
METROSERT	66	26,5	28	16
MIRS/UM-FS/LTM	70	18	50	20
NMI	248	31	20	24
DPM	32	28	12	24
LATMB	2	27	20	27
NPL	43	17	40	20
BFKH	43	20	40	30
METROSERT	57	26,5	14	16,5
NSAI NML	38	11	10	11
BMM	40	27	30	27
METROSERT	57	26,5	23	16

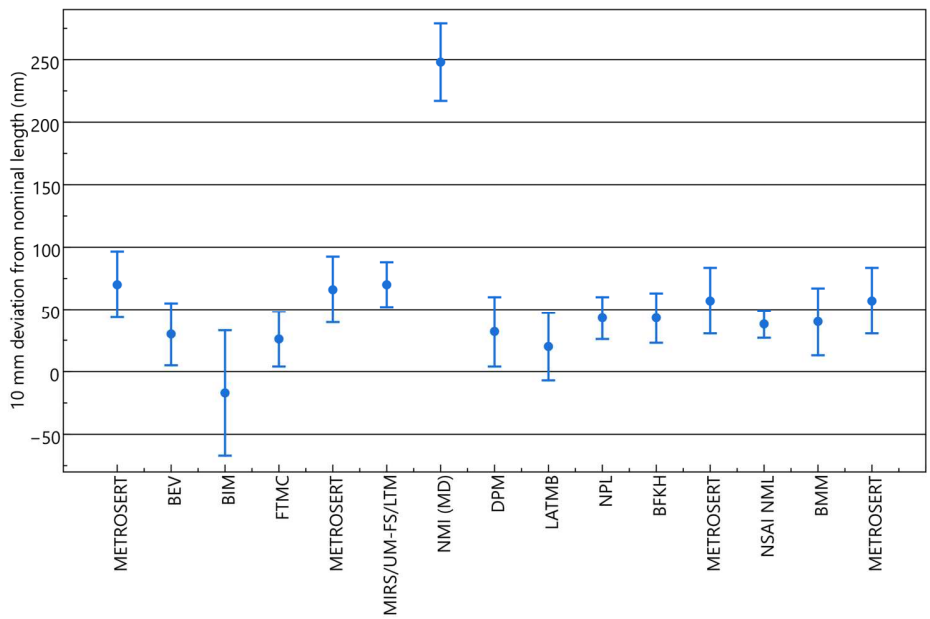

Figure 8. Deviation from nominal central length of the 10 mm steel gauge block (k = 1).

Table 7. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **25 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	29	28	28	16
BEV	-23	26	28	15
BIM	-17	52	50	57
FTMC	17	24	14	24
METROSERT	26	28	29	16
MIRS/UM-FS/LTM	10	19	40	20
NMI	220	32	30	24
DPM	-57	34	28	24
LATMB	10	29	30	29
NPL	46	19	27	20
BFKH	31	23	52	33
METROSERT	20	28	34	17
NSAI NML	6	14	30	14
BMM	10	33	30	33
METROSERT	26	28	30	17

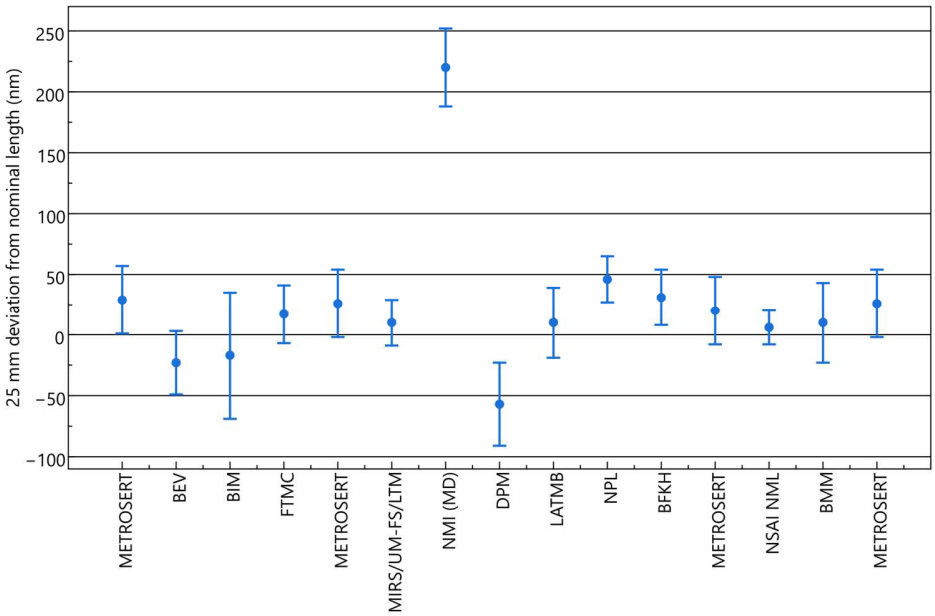

Figure 9. Deviation from nominal central length of the 25 mm steel gauge block (k = 1).

Table 8. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **50 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	88	36,5	38	16
BEV	37	27	54	15
BIM	24	56	40	61
FTMC	39	29	31	29
METROSERT	74	36,5	44	16,5
MIRS/UM-FS/LTM	70	22	60	20
NMI	63	33	30	24
DPM	9	47	47	24
LATMB	-30	35	50	35
NPL	120	25	47	20
BFKH	24	31	79	44
METROSERT	70	36,5	48	16,5
NSAI NML	50	21	20	21
BMM	50	48	60	48
METROSERT	67	36,5	51	17

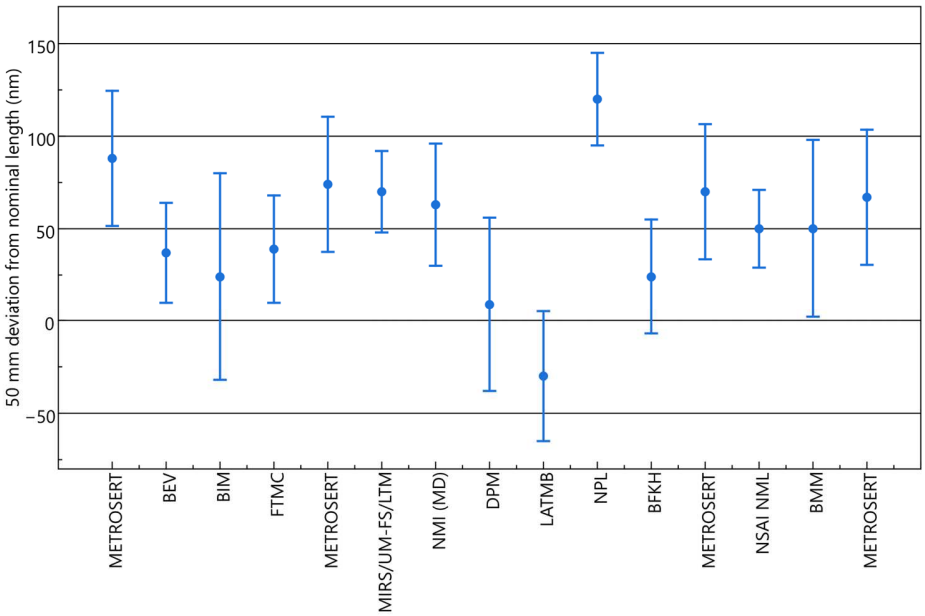

Figure 10. Deviation from nominal central length of the 50 mm steel gauge block (k = 1).

Table 9. Deviation from nominal central length, variation, and standard uncertainties (k = 1), as reported by laboratories in chronological order.

l_n **100 mm**	Δl_c, nm	$u(\Delta l_c)$, nm	v, nm	$u(v)$, nm
METROSERT	142	49	77	18
BEV	72	32	108	15
BIM	62	71	70	76
FTMC	116	44	67	44
METROSERT	105	49	101	2
MIRS/UM-FS/LTM	90	33	90	20
NMI	347	36	60	24
DPM	-245	81	46	24
LATMB	60	54	120	54
NPL	148	42	113	20
BFKH	100	60	121	95
METROSERT	117	49	88	185
NSAI NML	86	38	50	38
BMM	100	84	120	84
METROSERT	116	49	94	19

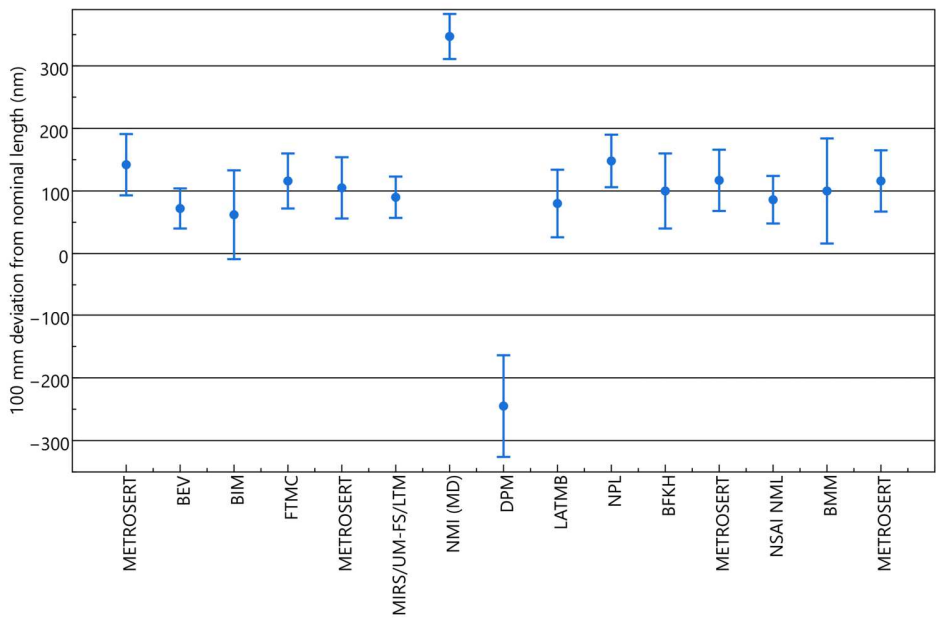

Figure 11. Deviation from nominal central length of the 100 mm steel gauge block (k = 1).

7.2. Measurement uncertainties

Table 10. Standard uncertainty components of deviation from nominal central length uncertainty budgets (u_c, k = 1; U_{CMC}, k = 2) provided by participants. L is l_n in mm. The entries indexed with superscript *a* signify an uncertainty estimation for a 100 mm nominal length gauge block.

	METROSERT	BEV	BIM	FTMC	MIRS/UM-FS/LTM	NMI	DPM	LATMB	NPL	BFKH	NSAI NML	BMM
$u(l_s)$	Q[10,5; 0,10 L]	15	0,15 L	14,5	Q[10; 0,09 L]	0,05	Q[13; 0,21 L]	11,5	Q[13; 0,21 L]	15,0	5,25	Q[17,5; 0,25 L]
$u(\theta)$	0,29 L	1,0	0,20 L	21,2	0,054 L		0,34 L		0,132 L	16,7	0,35	2,92E-4 L
$u(\theta_{a,m})$	0,087 L											
$u(\delta\theta)$	0,20 L	17	0,25 L	33,2	0,132 L	0,33	0,40 L		0,135 L	33,2	0,29	3,32E-4 L
$u(\alpha_s)$		0,12 L	0,17 L		0,030 L		0,035 L		0,116 L	2,89		5,77E-5 L
$u(\delta\alpha)$					0,021 L	0			0,231 L			4,08E-8 L
$u(\alpha_a)$		0,12 L	0,13 L				0,047 L			2,89		
$u(\alpha_s)u(\theta)$							0,117 L					
$u(\alpha_a)u(\theta)$							0,115 L	0,24E-6				
$u(\alpha_s)u(\delta\theta)$							0,027 L	0,24				
$u(c)$	15,1		30	15		21		17,3				
$u(d)$		1,5	20	4,02	Q[10,5; 0,19 L]	7,4	24		8,2			5,48
$u(d_s)$			12	2,25				6,9			3,8	16 + 6E-4 L
$u(\beta_{drift})$		0,144 L	0,25 L		0,082 L		0,115 L					0,102 L
$u(e_{drift})$	8,7	11,5		18,5				8,2		28,9	11,5	8,16
$u(\Delta p)$	5,0	5,0				9,62						4,04
$u(\varepsilon)$	0,192 L^2							2,9		2,89		
$u(\kappa)$	2				≅0			0,12E-6		7,22	5,7	
$u(\rho)$	20		20		4			10		32,7	3,0	
$u(\gamma)$	2,9									4,08		
u_c	Q[24; 0,43 L]	31,3[a]	Q[50; 0,5 L]	43,7[a]	Q[15; 0,25 L]	24,3	Q[27; 0,76 L]	Q[25,6; 0,47 L]	Q[16; 0,38 L]	60[a]	17.8	Q[26; 0,8 L]
U_{CMC}	Q[54; 0,91 L]	Q[50; 0,4 L]	Q[100; 1 L]	Q[48; 0,9 L]	Q[35; 0,5 L]	Q[50; 0,84 L]	N/A	Q[80; 1,3 L]	Q[32; 0,76 L]	Q[40; 1,3 L]	Q[83; 22,32 L]	[54 + 1,2 L]

Uncertainty component designations:

$u(l_s)$ – standard gauge block uncertainty

$u(\theta)$ – temperature deviation of standard gauge block from reference temperature uncertainty

$u(\theta_{a,m})$ – air or material temperature measurement uncertainty

$u(\delta\theta)$ – temperature difference between gauge blocks uncertainty

$u(\alpha_s)$ – coefficient of thermal expansion of standard uncertainty

$u(\alpha_a)$ – coefficient of thermal expansion of artifact uncertainty

$u(\delta\alpha)$ – difference of coefficient of thermal expansion uncertainty

$u(d)$ – length difference of gauge blocks uncertainty

$u(d_s)$ – surface roughness of artefact gauge block uncertainty

$u(e_{drift})$ – constant component of drift of standard gauge block deviation from nominal central length uncertainty

$u(\beta_{drift})$ – length dependent component of drift of standard gauge block deviation from nominal central length uncertainty

$u(\Delta p)$ – flatness difference between standard and artifact gauge block uncertainty

$u(\varepsilon)$ – deformation uncertainty

$u(c)$ – comparator uncertainty

$u(\kappa)$ – comparator linearity uncertainty

$u(\rho)$ – comparator repeatability uncertainty

$u(\gamma)$ – resolution of consecutive readings uncertainty

8 Analysis

8.1. Calculation of the SCRV, Degrees of Equivalence, and Chi-squared

Regarding correlation, the source of traceability for interferometric calibration of reference gauge blocks was GUM for three participants (Metrosert, FTMC, NSAI NML), and VTT MIKES for two participants (BIM, LATMB). The traceability source was unique for all other participants. Thus, the reported standard gauge block uncertainty components for those participants were used for covariance matrix calculations when calculating the reference value (RV) for the Deviation from nominal central length.

Outliers were eliminated iteratively on a per-gauge and measurand basis before calculating the reference value by eliminating the participant with the largest E_n magnitude until the Birge ratio satisfied a value of $R_B < \sqrt{1+\sqrt{\frac{8}{I-1}}}$, where I is the number of contributing laboratories, when considering standard uncertainties (k = 1).

The Birge ratio R_B is derived from the application of least squares algorithms and the χ^2-test:

$$R_B = \frac{u_{ext}(\bar{x}_w)}{u(\bar{x}_w)}$$

where $u_{ext}(\bar{x}_w)$ is the external standard deviation

$$u_{ext}(\bar{x}_w) = \sqrt{\frac{1}{(I-1)} \cdot \frac{\sum_{i=1}^{I} w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^{I} w_i}},$$

w_i is the normalised weight for the result x_i

$$w_i = C \cdot \frac{1}{[u(x_i)]^2},$$

and the normalising factor, C, is given by

$$C = \frac{1}{\sum_{i=1}^{I} \left[\frac{1}{u(x_i)}\right]^2}.$$

The weighted mean, $\bar{x}_w$, is calculated by:

$$\bar{x}_w = \sum_{i=1}^{I} w_i \cdot x_i$$

The uncertainty of the weighted mean is calculated by:

$$u(\bar{x}_w) = \sqrt{\frac{1}{\sum_{i=1}^{I} \left[\frac{1}{u(x_i)}\right]^2}} = \sqrt{C}$$

The equation for the R_B cutoff is derived from the Birge ratio squared, which is effectively the reduced chi-squared statistic (χ_υ^2/υ), where the degrees of freedom is $\upsilon = I - 1$ with I being the number of contributing laboratories.

From standard statistics, the properties of a chi-square distribution dictate:

Expected value: $E[\chi^2] = \upsilon$

Variance: $Var(\chi^2) = 2\upsilon$

To find the properties of the reduced chi-squared (R_B^2):

Expected value of R_B^2: $\frac{\upsilon}{\upsilon} = 1$

Variance of R_B^2: $\frac{2\upsilon}{\upsilon^2} = \frac{2}{\upsilon} = \frac{2}{I-1}$

Standard deviation of R_B^2: $\sigma = \sqrt{\frac{2}{I-1}}$

To establish a threshold for 95 % confidence (k = 2), two standard deviations are added to the expected value:

$$R_B^2 < 1 + 2\sigma \Leftrightarrow R_B^2 < 1 + 2\sqrt{\frac{2}{I-1}} \Leftrightarrow R_B^2 < 1 + \sqrt{\frac{8}{I-1}} \Leftrightarrow R_B < \sqrt{1 + \sqrt{\frac{8}{I-1}}}$$

After deriving the weighted mean and its associated standard uncertainty, the deviation of each laboratory's result from the weighted mean is determined as $x_i - \bar{x}_w$. The uncertainty of this deviation is calculated as a combination of the uncertainties of the result, $u(x_i)$, and the uncertainty of the weighted mean $u(\bar{x}_w)$. The uncertainty of the deviation from the weighted mean, which includes the covariance term for laboratories with a correlated source of traceability, and excludes the covariance term for other participants, is given by:

$$u(x_i - \bar{x}_w) = \sqrt{[u(x_i)]^2 - [u_{int}(\bar{x}_w)]^2}$$

The degree of equivalence (DoE) is calculated as:

$$DoE_i = x_i - \bar{x}_w$$

The uncertainty of DoE is calculated depending on whether the participant was an outlier [3].

For included participants:

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2 - 2 \cdot cov(x_i; \bar{x}_w)}$$

For excluded participants:

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2}$$

Note that the equality $cov(x_i; \bar{x}_w) = u^2(\bar{x}_w)$ breaks down because the participants are neither perfectly independent nor perfectly dependent due to shared traceability chains between sets of

some but not all laboratories. Thus, as $cov(x_i; \bar{x}_w) < u^2(\bar{x}_w)$, for included participants with less than identical uncertainty traceability:

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2 - 2 \cdot cov(x_i; \bar{x}_w)} \neq \sqrt{[u(x_i)]^2 - [u_{int}(\bar{x}_w)]^2}$$

E_n, which is used to check for statistical consistency of the results with their associated uncertainties, is defined as the ratio of the deviation from the weighted mean, divided by the expanded uncertainty of this deviation:

$$E_n = \frac{x_i - \bar{x}_w}{2 \cdot u_{DoE,i}}$$

The covariance of x_i with $\bar{x}_w$ is calculated by:

$$cov(x_i; \bar{x}_w) = \sum_{i=1}^{I} w_i V_{ij}$$

where the covariance matrix *V* is constructed as:

$$V_{ij} = \begin{cases} u_i^2 & \text{if } i = j \\ u_{common,i} \cdot u_{common,j} & \text{if } i \neq j \end{cases}$$

$u_{common,i}$ being the uncertainty contribution within u_i originating from the same source of traceability as $u_{common,j}$ in u_j.

While the Birge ratio evaluates the consistency of the consensus group for a single measurand, judging whether an individual participant's results support their overall uncertainty claim requires evaluating their performance across all measured artifacts. Because comparisons involving several measurands increase the statistical risk that a participant with correctly evaluated uncertainties may still observe an isolated large normalized error, a systematic acceptance procedure is required.

Consider a single participant reporting results for different measurands relating to a particular Calibration and Measurement Capability (CMC) claim. The Degree of Equivalence (DoE) for the *i*-th measurand is denoted as DoE_i, representing the difference between the participant's result and the SCRV, alongside its associated standard uncertainty $u(DoE_i)$. The uncertainty accounts for the uncertainty of the SCRV, the participant's claimed uncertainty, and the mathematical correlation between the two if the participant contributed to the reference value.

The assessment proceeds in two stages:

Normalized Error Criterion

First, the normalized errors are computed as $E_{n,j} = DoE_i / \left(k \cdot u(DoE_i)\right)$, utilizing a coverage factor of *k* = 2 for approximately 95% confidence. If $|E_{n,i}| \leq 1$ for all of the participant's

measurements, indicating the difference from the SCRV is less than the expanded uncertainty of this difference across all artifacts, the participant's uncertainty claim is immediately accepted.

Multivariate Chi-Squared Assessment

If the participant observes $|E_{n,i}| > 1$ for one or more measurements, the chi-squared test statistic is computed to evaluate their overall capability. Because a single laboratory's measurements of multiple gauge blocks share systematic error components (e.g., reference standard traceability, assumed thermal expansion coefficients), the measurands are not strictly independent. Briefly, the Chi-squared test statistic was calculated for participants that failed the E_n and Birge ratio test per WG-MRA GD-3 v1.7 chapter 5.2.

To account for this intra-participant correlation, the test statistic is calculated using the Mahalanobis distance squared:

$$\chi^2_{obs} = (DoE)^{\mathrm{T}} cov^{-1} DoE = d^{\mathrm{T}} C^{-1} d$$

d – the column vector of the participant's DoE values

d^{T} – the transpose of the DoE vector

C – intra-participant covariance matrix of the DoEs

The elements of the covariance matrix *C* are constructed from the participant's standard uncertainty budget. The diagonal elements represent the total variance of the difference for each individual block:

$$C_{ii} = u^2(DoE_i)$$

The off-diagonal elements represent the covariance between the *j*-th and *k*-th blocks, driven by the shared fixed standard uncertainty (u_f) and the length-dependent standard uncertainty coefficient (u_l) applied to lengths L_j and L_k:

$$C_{jk} = u_f^2 + (u_l \cdot L_j)(u_l \cdot L_k)$$

Decision Rule The resulting test statistic χ^2_{obs} is compared to $\chi^2_{95,m}$, which is the 95th percentile of the chi-square distribution with *m* degrees of freedom.

- If $\chi^2_{obs} < \chi^2_{95,m}$, the participant's overall uncertainty claim is accepted, and the isolated values are deemed statistical anomalies.
- If $\chi^2_{obs} > \chi^2_{95,m}$, the threshold is exceeded and there is a systemic problem; the participant's results and uncertainty claim must be investigated and formally flagged.

The measured and reference values for the two measurands are listed in Table 11 and Table 12 with the associated standard uncertainty and Birge ratio. The gauge specific values for E_n, w, DoE and u_{DoE} for all participants are listed in tables Table 13 through Table 18.

Table 11. Measured values and time-average of pilot, and reference value $\overline{\Delta lc_w}$ of gauge block deviation from nominal central length, associated standard uncertainty $u(\overline{\Delta lc_w})$, Birge ratio, and chi-squared. The Birge ratio represents the value obtained after excluding outlier and disqualified participants per gauge block.

		Gauge block nominal length (mm)					
Meas. date		0,5	4	10	25	50	100
2023.07.23	Δlc_w(nm)	37	46	70	29	88	142
	$u(\Delta lc_w)$(nm)	23,5	23,5	26,5	28	36,5	49
2024.09.30	Δlc_w(nm)	34	36	66	26	74	105
	$u(\Delta lc_w)$(nm)	23,5	23,5	26,5	28	36,5	49
2024.01.15	Δlc_w(nm)	45	27	57	20	70	117
	$u(\Delta lc_w)$(nm)	23,5	23,5	26,5	28	36,5	49
2025.04.08	Δlc_w(nm)	41	33	57	26	67	116
	$u(\Delta lc_w)$(nm)	24	23,5	26,5	28	36,5	49
$\Delta lc_{w,t-avg}$ (nm)		38,834	35,375	62,681	25,255	74,474	118,357
$u(\Delta lc_{w,t-avg})$ (nm)		24	24	27	28	37	49
$\overline{\Delta lc_w}$ (nm)		30,4	35,6	41,6	14,6	54,4	94,9
$u(\overline{\Delta lc_w})$ (nm)		6,2	6,3	6,7	7,6	9,2	14,4
R_B		1,017	1,084	0,834	0,845	1,248	0,600
I		11	11	10	10	11	10
R_B cutoff		1,377	1,377	1,394	1,394	1,377	1,394
Participants excluded		DPM, NMI	DPM, NPL	DPM, NMI	DPM, NMI	DPM, LATMB, NPL	DPM, NMI

Table 12. Measured values and time-average of pilot, and reference value $\bar{v}_w$ of gauge block variation, associated standard uncertainty $u(\bar{v}_w)$, Birge ratio, and chi-squared. The Birge ratio represents the value obtained after excluding outlier and disqualified participants per gauge block.

		Gauge block nominal length (mm)					
Meas. date		0,5	4	10	25	50	100
2023.07.23	v_w (nm)	9	23	18	28	38	77
	$u(v_w)$ (nm)	15,5	16	15,5	16	16	18
2024.09.30	v_w (nm)	20	27	28	29	44	101
	$u(v_w)$ (nm)	16	16	16	16	16,5	20
2024.01.15	v_w (nm)	16	23	14	34	48	88
	$u(v_w)$ (nm)	16	16	16,5	17	16,5	18,5
2025.04.08	v_w (nm)	25	32	23	30	51	94
	$u(v_w)$ (nm)	16	17	16	17	17	19
$\Delta v_{w,t-avg}$ (nm)		17,851	26,213	21,620	30,076	45,061	90,495
$u(v_{w,t-avg})$ (nm)		16	17	17	17	17	20
$\bar{v}_w$ (nm)		21,6	28,4	24,2	30,2	45,7	89,2
$u(\bar{v}_w)$ (nm)		5,5	5,6	5,7	6,2	6,7	7,6
$R_{B,v}$		0,496	0,535	0,725	0,345	0,580	0,938
I		12	12	12	12	12	12
R_B cutoff		1,361	1,361	1,361	1,361	1,361	1,361
Participants excluded		None	None	None	None	None	None

Table 13. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation from nominal central length and variation for 0,5 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,182	0,029	9	24	0,123	0,118	-4	16
BEV	0,195	0,062	-10	25	0,270	0,134	-8	14
BIM	0,026	0,016	3	50	0,169	0,010	19	55
FTMC	0,346	0,015	15	22	0,387	0,063	17	22
MIRS/UM-FS/LTM	0,013	0,119	-1	17	0,221	0,076	9	20
NMI	1,049	0,043	62	30	0,247	0,053	-12	24
DPM	0,369	0	-21	28	0,290	0,053	-14	24
LATMB	0,405	0,057	-21	26	0,031	0,045	-2	26
NPL	0,766	0,151	23	15	0,221	0,076	9	20
BFKH	0,222	0,097	-9	20	0,451	0,032	28	31
NSAI NML	0,793	0,359	-13	8	0,092	0,302	-2	9
BMM	0,190	0,057	10	26	0,031	0,045	-2	26

Table 14. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation from nominal central length and variation for 4 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,005	0,029	-1	24	0,066	0,106	-3	17
BEV	0,467	0,063	-23	25	0,132	0,135	4	14
BIM	0,379	0,016	-38	50	0,016	0,011	2	55
FTMC	0,488	0,015	-21	22	0,196	0,063	-9	22
MIRS/UM-FS/LTM	0,166	0,120	-6	17	0,564	0,076	22	20
NMI	0,28	0,044	17	30	0,179	0,053	-9	24
DPM	0,336	0	-19	28	0,008	0,053	-1	24
LATMB	0,507	0,058	-26	26	0,164	0,045	-9	26
NPL	1,151	0,135	37	16	0,382	0,076	15	20
BFKH	0,825	0,108	-30	18	0,153	0,037	9	29
NSAI NML	0,410	0,361	7	8	0,500	0,304	-9	9
BMM	0,111	0,058	-6	26	0,230	0,045	12	26

Table 15. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation from nominal central length and variation for 10 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,404	0,03	22	27	0,079	0,113	-3	17
BEV	0,24	0,071	-12	25	0,463	0,144	13	14
BIM	0,591	0,018	-59	50	0,054	0,011	6	55
FTMC	0,372	0,03	-16	21	0,333	0,067	-15	22
MIRS/UM-FS/LTM	0,850	0,137	29	17	0,675	0,081	26	20
NMI	3,256	0	207	32	0,089	0,057	-5	24
DPM	0,375	0	-22	29	0,261	0,057	-13	24
LATMB	0,412	0,061	-22	27	0,079	0,045	-5	27
NPL	0,046	0,153	2	16	0,414	0,081	16	20
BFKH	0,039	0,111	2	19	0,27	0,036	16	30
NSAI NML	0,204	0,334	-4	9	0,752	0,268	-15	10
BMM	0,030	0,061	-2	27	0,111	0,045	6	27

Table 16. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation from nominal central length and variation for 25 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,197	0,044	11	27	0,003	0,131	-1	16
BEV	0,757	0,085	-38	25	0,079	0,168	-3	14
BIM	0,308	0,021	-32	52	0,176	0,012	20	57
FTMC	0,052	0,054	3	23	0,349	0,066	-17	24
MIRS/UM-FS/LTM	0,133	0,159	-5	18	0,259	0,095	10	20
NMI	3,123	0	206	33	0,004	0,066	-1	24
DPM	0,598	0	-42	35	0,047	0,066	-3	24
LATMB	0,083	0,067	-5	28	0,003	0,045	-1	29
NPL	0,900	0,159	32	18	0,083	0,095	-4	20
BFKH	0,377	0,108	17	22	0,337	0,035	22	33
NSAI NML	0,367	0,259	-9	12	0,006	0,193	-1	13
BMM	0,070	0,050	-5	34	0,003	0,035	-1	33

Table 17. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation of central length and variation for 50 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,285	0,045	21	36	0,019	0,176	-1	15
BEV	0,342	0,115	-18	26	0,313	0,200	9	14
BIM	0,275	0,025	-31	56	0,047	0,013	-6	61
FTMC	0,279	0,075	-16	28	0,259	0,054	-15	29
MIRS/UM-FS/LTM	0,391	0,173	16	21	0,382	0,113	15	19
NMI	0,137	0,077	9	32	0,339	0,078	-16	24
DPM	0,028	0	3	48	0,031	0,078	2	24
LATMB	1,249	0,067	-85	34	0,064	0,037	5	35
NPL	1,411	0,134	66	24	0,037	0,113	2	19
BFKH	0,513	0,087	-31	30	0,384	0,024	34	44
NSAI NML	0,116	0,171	-5	19	0,644	0,102	-26	20
BMM	0,047	0,037	-5	48	0,152	0,020	15	48

Table 18. E_n, weight, degree of equivalence and associated standard uncertainties (k = 1) for deviation from nominal central length and variation for 100 mm gauge block.

	Deviation from nominal central length				Variation			
Participant	E_n	w	DoE	u_{DoE}	E_n	w	DoE	u_{DoE}
METROSERT	0,251	0,074	24	47	0,037	0,144	2	19
BEV	0,4	0,201	-23	29	0,729	0,256	19	13
BIM	0,237	0,039	-33	70	0,127	0,01	-20	76
FTMC	0,255	0,094	22	42	0,256	0,03	-23	44
MIRS/UM-FS/LTM	0,082	0,189	-5	30	0,023	0,144	1	19
NMI	3,254	0	253	39	0,641	0,1	-30	23
DPM	0,565	0	-93	83	0,948	0,1	-44	23
LATMB	0,335	0,069	-35	53	0,289	0,02	31	54
NPL	0,674	0,117	54	40	0,645	0,144	24	19
BFKH	0,045	0,058	6	59	0,169	0,007	32	95
NSAI NML	0,126	0,134	-9	36	0,526	0,04	-40	38
BMM	0,032	0,03	6	83	0,185	0,009	31	84

Table 19. Maximum E_n, chi-squared statistics, and assessment status of participants.

Participant	Deviation from nominal central length				Variation			
	$E_{n,max}$	χ^2	χ^2_{95}	Assessment	$E_{n,max}$	χ^2	χ^2_{95}	Assessment
METROSERT	0,41	-9,08	12,59	Accept $E_n \leq 1$	0,13	0,15	12,59	Accept $E_n \leq 1$
BEV	0,76	2,72	12,59	Accept $E_n \leq 1$	0,73	5,62	12,59	Accept $E_n \leq 1$
BIM	0,6	-38,54	12,59	Accept $E_n \leq 1$	0,18	0,58	12,59	Accept $E_n \leq 1$
FTMC	0,49	4,85	12,59	Accept $E_n \leq 1$	0,39	3,27	12,59	Accept $E_n \leq 1$
MIRS/UM-FS/LTM	0,85	23,96	12,59	Accept $E_n \leq 1$	0,68	3,13	12,59	Accept $E_n \leq 1$
NMI	3,26	24,68	12,59	Investigate	0,65	2,45	12,59	Accept $E_n \leq 1$
DPM	0,60	20,77	12,59	Accept $E_n \leq 1$	0,95	5,78	12,59	Accept $E_n \leq 1$
LATMB	1,25	7,93	12,59	Accept χ^2 passed	0,29	0,97	12,59	Accept $E_n \leq 1$
NPL	1,42	-0,22	12,59	Accept χ^2 passed	0,65	3,34	12,59	Accept $E_n \leq 1$
BFKH	0,83	14,99	12,59	Accept $E_n \leq 1$	0,46	1,16	12,59	Accept $E_n \leq 1$
NSAI NML	0,80	6,38	12,59	Accept $E_n \leq 1$	0,76	4,87	12,59	Accept $E_n \leq 1$
BMM[a]	0,19	-1,05	12,59	Accept $E_n \leq 1$	0,23	0,51	12,59	Accept $E_n \leq 1$

[a]Updated values were assessed.

8.2. Participant responses

DPM (D. Bulku, September 17, 2025):

First of all, we would like to thank you for Report Draft A. As noted in the report, our laboratory had two outlier results at the lengths of 25 mm and 100 mm.

We would like to inform you that, following an internal review and analysis of the comparison results, we identified a **technical error in the Excel data processing**, which affected the results we initially reported. We understand that it is now too late to make formal corrections.

During our review, we clearly identified that not only were there two outlier results, but due to this technical error, **all calibration results from our laboratory were not traceable to the EIM reference values**.

Instead of using the correct traceable central length formula **Lc = (Lref + R2 - R1)**, the Excel cell mistakenly used **Lc = Ln + R2 - R1**, which caused the results to fall outside of traceability.

Furthermore, we incorrectly reported the comparison measurement difference (R2 - R1) as the deviation in central length, that is completely different to real traceable central length deviation.

Following the correction of the central length deviation values, according to the Report Draft A , all comparison results seem to be statistically significant.

DPM (D. Bulku, September 18, 2025):

I wrote you before, about the mistake in calculation of comparison results. So, for more info please find in attach file the doc:

- Appendix C 2024
- Appendix C reviewed 2025
- Excel doc. explaining the case of error.

As the results are published, Draft A, there is no possibility changing anything. Well, I stated also that, other results are incorrect, even statistically they seem to be significant. Does the report address it in a way?

BMM (T. Božović, August 22, 2025)

Thank you for your email and the report draft A1.

I would like to thank all participants for the opportunity to repeat the measurements and I apologize as it caused an extension of the entire process.

There is a mistake regarding our measurement uncertainty data in the Draft A1.

In our Appendix C are given u(lc) and u(v) not U(lc) and U(v), so the values shouldn’t be divided by 2.

Our mistake is that we reported a difference of measurement uncertainty in relation to our current CMC, as significant.

It's similar with the reported damages on the standards. Maybe we shouldn't have reported them as significant, except for the reference of 100 mm.

I apologize for that because it caused additional work in processing the measurement results.

A lesson has been learned for next time.

NPL (A. Lewis, September 2, 2025):

I note the two NPL results which do not agree with the reference value within the uncertainties.

We are looking into the measurement data for these and will give some possible explanations after an internal discussion. Sorry that we were such a bad linking lab for these two gauges!

I am also looking through the report in more detail but this will take some time as I have a mountain of other things to do.

I noticed a few small errors in the results tables where the METROSERT uncertainty sometimes jumps by x10 (due to missing decimal separator). I assume these are just in the display and not in the actual numbers in the analysis.

Three items to note in the analysis:

1. Only 1 METROSERT result can be included for each gauge – it should be the first measurement – the one you did before you saw any other results.
2. For the two NPL 'bad' results, you have excluded them from the calculations and from the results display. Our usual process is to only exclude them when the Birge ratio check is not satisfied. Then we start excluding results (set the weighting to zero) starting with the result with the highest En value, and stop excluding results from the calculation when the Birge ratio test is OK. I think this means that the NPL results for the 4 mm and 50 mm gauges should sill be part of the calculation, because the Birge ratio test is OK despite the bad NPL results.
3. Even laboratories with bad results should have their En values in the tables, as well as DoE values.

There is another statistical test we can perform across the set of all NPL results to see if there is a big problem or not, but I have nt yet had time to do this. It is a process proosed by Brian Eves at a recent WG-MRA meeting that we plan to implement in the guidance. When I have a chance to try it, I will share with you.

METROSERT reply (J. S. Eensalu, September 3, 2025)

Thank you for your feedback

1. In the analysis, the time-averaged Metrosert result was used for each gauge block as per chapter 7.2 in the technical protocol: Steel gauge blocks occasionally show a

growing or a shrinkage the rate of which is approximately linear with time. Since the artefacts used here are new and with no history, the instability of the blocks must be determined in course of the comparison. For this check the measurements of the pilot laboratory are used exclusively, not that of the other participants. Using these data a linear regression line is fitted and the slope together with its uncertainty is determined (per gauge block). Three cases can be foreseen: a) The linear regression line is an acceptable drift model and the absolute drift is smaller than its uncertainty. The gauge block is considered stable and no modification to the standard evaluation procedure will be applied. In fact the results of the pilot's stability measurements will not influence the numerical results in any way.

The case A applies in this comparison based on our measurement results (see Figures 4 and 5 in the report).

2. b) The linear regression line is an acceptable drift model and the absolute drift is larger than its uncertainty, i.e. there is a significant drift for the gauge block. In this case an analysis similar to [Nien F Z et al. 2004, Statistical analysis of key comparisons with linear trends, Metrologia 41, 231] will be followed. The pilot influences the reference value by the slope of the drift only, not by the measured absolute lengths. The data are not compatible at all with a linear drift, regarding the uncertainties of the pilot's measurements. In this case the artefact is unpredictably unstable or the pilot has problems with its measurements. TC-L has to determine the further approach.

3. Unfortunately, the results of these two gauge blocks did not satisfy the Birge ratio requirement of RB < 1 when NPL was included. I followed the procedure of first including all participants, and thereafter eliminating the participant with the highest En magnitude one by one iteratively until the Birge ratio decreased below 1. The Birge ratio presented is the value obtained after eliminating the mentioned participants. Perhaps this fact should be expressed more explicitly in the report. Based on the statistical analysis all outliers by all outlined participants were significant and none were in the grey zone. It is possible that most participants are not taking into consideration something that NPL does, which would explain the somewhat positive bias of the central length deviation for NPL for all gauge blocks, but without further tests it is difficult to draw conclusions as to what this originates from.

4. I will add the En results of all participants in the tables. DoE is already included for all participants. The minus sign meant that that participant was not included in the weighting.

5. The uncertainty values for Metrosert in the tables were indeed missing decimals in a few places. I appreciate that you noticed. This affected only the table results, however, as the time-averaged values were used in the analysis.

NPL (A. Lewis, September 26, 2025)

Sorry for the delay in responding – too much work to do!

I agree with using the time-weighted METROSERT result for each gauge block. This should be the only METROSERT result in the En calculations. It would be good to show these time-weighted results somewhere. I suggest adding the time-weighted values to the tables of results. This will help anyone who is checking the tables perform their own En calculations using on the time-weighted result for METROSERT. I tried to do this but I get slightly different En values to the ones you report.

For example:

Using 35.5 nm as the average result for METROSERT for the 4 mm gauge, I get a Birge ratio that is OK with all participants included. The NPL En value is 1.17.

Using 62.5 nm for METROSERT for the 10 mm gauge, the Birge ratio fails until we exclude NMI (MD)

Using 25.3 nm for METROSERT for the 25 mm gauge, the Birge ratio fails until we exclude NMI (MD) (we do not need to exclude DPM).

Using 74.8 nm for METROSERT for the 50 mm gauge, the Birge ratio is OK with all included. NPL En is 1.4 and LATMB En is 1.21.

I attach my analysis Excel.

We have the following comments on the discrepant NPL results.

4 mm gauge block

The NPL master gauge was calibrated 3/2022.

Master gauge re-calibration in 09/2025 shows zero change in length.

Gauge showed pitting and light scratches which cross the measurement region.

Our 3 results had a 10 nm spread.

50 mm gauge block

The NPL master gauge was calibrated 3/2022.

Master gauge re-calibration in 09/2025 shows +11 nm change in length. That would make our result longer

Gauge had some pitting but not in the measurement region.

Our 3 results had a 10 nm spread.

So, nothing obvious as to the source of the problem. Our comparator is very old (40+ years) and starting to show issues which affect its zeroing. It is possible this is the cause of the problem. Normally comparator results agree with the GBI results within the uncertainties (i.e. within 40 nm) but as the protocol did not allow interferometry measurements, we did not put the gauges in the GBI.

METROSERT reply (J. S. Eensalu, October 1, 2025):

I looked at the calculations you sent. I admit not presenting the time averaged values makes it nigh impossible for others to reproduce the results obtained. The Python scripts I used to analyse the data are attached with the output Excel files.

The time averaged Metrosert central length deviation values used in our respective analyses were quite a bit different. Hence the differences in results. I am intrigued as to whether our results will match when using the same initial values from A1.

Note that I excluded participants if the Birge ratio was even slightly above 1.000.

Meanwhile, DPM has informed us that all of their central length deviation results were miscalculated due to using the nominal value instead of the central length deviation of their standard.

After consultation with Dariusz, all results by DPM are to be excluded from the RV calculation, and I will have to analyze their results separately in an additional table and present relevant graphs in A2, passing a decision based on their updated data, as Tim had previously suggested to Dariusz for another supplementary comparison.

Perhaps it would be prudent to suggest at the upcoming TC-L that future gauge block supplementary comparisons should make use of interferometric measurements by the pilot or co-pilot to mitigate the emergence of such complicated conundrums as much as possible? What is your opinion on this matter?

NPL (A. Lewis, 14 October, 2025)

Nice to see you over the last 2 days (even if only on a screen).

Just a quick note to confirm that the TC-L meeting recommended to not use interferometric references for comparisons such as this due to the measurand being slightly different. Also, Antti Lassila mentioned to Tim during lunch that supplementary comparisons do not need the linking that it used in key comparisons.

So the mechanical results should stand on their own and it looks like NPL has some issues with their comparator

NPL (A. Lewis, 02 February, 2026)

Thanks for the update. It’s never easy being a pilot!

Tim has now taken over from me as the TC-L Delegate but he’s not the technical expert for the gauge block area so I am happy to respond with the following observations.

1. Section 7.2 caption for table 10 : italicise the k in U_{CMC}, k = 2.
2. Table 11 – for the 10 mm gauge block, NMI was also excluded (as shown in Table 15).
3. Table 11 – for the 100 mm gauge block, NMI was also excluded (as shown in Table 18).
4. For the results tables where some participants have been excluded (weight = 0), it is customary to still include their En value, (even though they are not contributing to the weighted mean). This is necessary for discussion of whether or not their results support the CMC claim. Tables 13 to 18 should include the (>1) En values for the labs that were excluded from the weighted mean.

5. In section 8.1 you state that participants are eliminated, in order, until the Birge ratio is less than 1. According to the CCL guidance document GD-3.2 (the report template) the Birge ratio has an expectation value of 1 only when considering standard uncertainties. For a coverage factor of *k* = 2 the expectation value is increased and the data in a comparison are consistent provided that

$$R_B < \sqrt{1 + \sqrt{8/(I-1)}}$$

where I is the number of laboratories contributing data. For this comparison RB is around 1.4 due to there being between 8 and 10 contributions to each weighted mean. I am not a statistics expert so I do not recall where this expectation value comes from but it has been used in all the comparison reports and is embedded in the spreadsheet used by several pilots. I analysed the results using this expectation value and in several cases the Birge ratio is satisfied without having to eliminate anyone. I should clarify that I was using the raw uncertainty data without taking into account correlations. My initial check suggests that the overall change is negligible – those with high En values still remain high and there is no change to the overall outcome.

6. The uncertainty formula for U(DoE) for included participants is stated as

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2 - 2 \cdot cov(x_i; \bar{x}_w)}$$

In situations where there are no correlations, the formula stated in the guidance document for results not contributing to the weighted mean is

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 - [u_{int}(\bar{x}_w)]^2}$$

So I wonder if the first sign in your equation should be a '-', as when there is no correlation, the term $2 \cdot cov(x_i; \bar{x}_w)$ is zero ? [J.S.E 12 March 2026: see updated chapter 8.1 on why I disagree]

7. I don't think the corrected analysis benefits the two NPL results with En>1. In both cases we measure 'long'. Since the comparison, we have re-calibrated most of the gauge blocks we used as references. The results for the 4 mm gauge are unchanged. The results for the 50 mm gauge are now 11 nm shorter, but we still get En ~ 1.1.

8. Showing the NPL En values for all gauges (and for all other participants who were excluded) is useful because the CCL has now issued guidance on analysing multiple artefact comparisons, to decide if there is support for CMC claims – see section 5.2 of the GD3 document (v1.7). This requires chi-squared test statistics to be performed using the calculated En values.

Sorry to not have spotted these potential issues sooner.

8.3. Analysis of revised DPM results

According to calculated values in Table 20, illustrated in Figure 12, the revised deviation from nominal central length of DPM supports the uncertainty claim based on chi-squared analysis. Thus, further corrective actions are not necessary. The corrected values have a systematic net negative bias vs the SCRV.

Table 20. Comparison of SCRV to revised DPM values for deviation from nominal central length.

	Gauge block nominal length (mm)					
	0,5	4	10	25	50	100
$\overline{\Delta lc_w}$ (nm)	30,4	35,6	41,6	14,6	54,4	94,9
$u(\overline{\Delta lc_w})$ (nm)	6,2	6,3	6,7	7,6	9,2	14,4
Δlc_{DPM} (nm)	10	17	20	-27	57	2
$u(\Delta lc_{DPM})$ (nm)	26	26	27	33	48	84
E_n	0,37	0,34	0,38	0,60	0,03	0,57
χ^2	5,78					
χ^2_{95}	12,59					

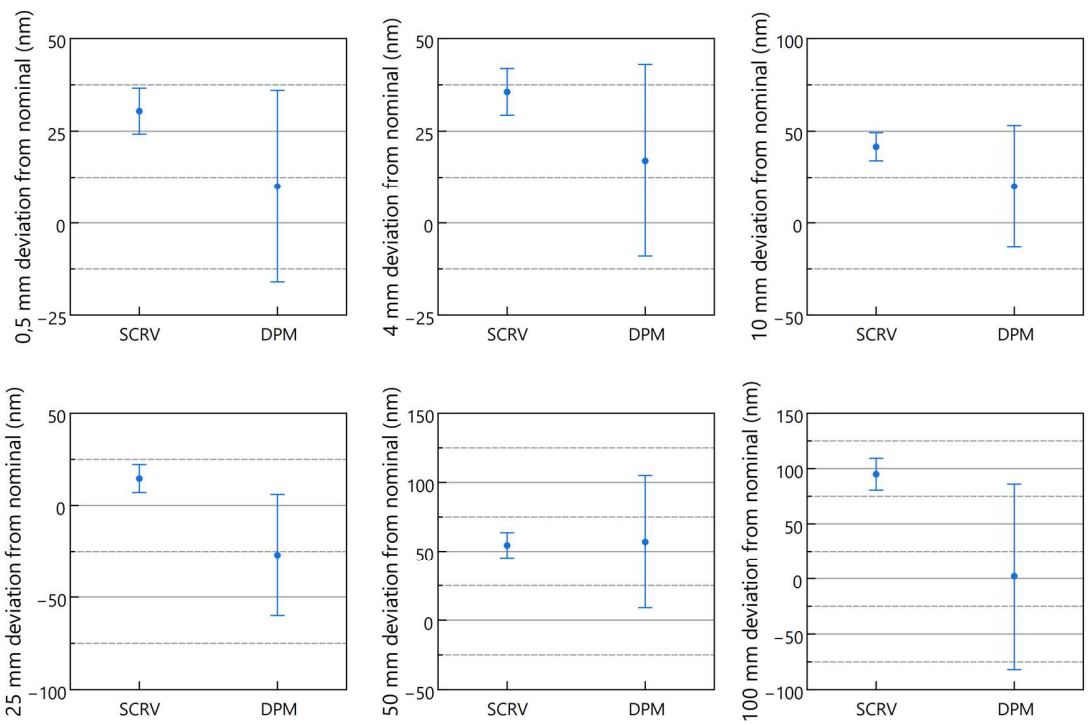

Figure 12. Comparison of SCRV to revised DPM values for deviation from nominal central length.

8.4. Discussion of results

Outliers were eliminated per measurand per gauge block until the Birge ratio was less than the specified value. Weights were not capped for analysis.

The statistical data analysis shows that NMI was an outlier for the 0,5 mm, 10 mm, 25 mm, and 100 mm gauge block deviation from nominal central length. There was a systematic net positive bias in their deviation from nominal central length measurement procedure.

NPL was an outlier for the deviation from nominal central length of the 4 mm and 50 mm gauge blocks in the E_n analysis, also exhibiting a net positive bias in the deviation from nominal central length measurement procedure. According to chi-squared analysis the CMC is supported and further corrective actions are not required.

Based on the originally submitted data, DPM was an outlier for the deviation from nominal central length of the 25 mm and the 100 mm gauge blocks with a net negative bias in general in comparison to the reference value.

LATMB was an outlier for the deviation from nominal central length of the 50 mm gauge block with net negative bias. Chi-squared analysis indicates the CMC claim is supported and further corrective actions are not needed.

NMI, NPL and LATMB were asked without disclosing results whether they wanted to adjust their reported results. NMI and NPL chose not to change their results. LATMB provided updated data for the 100 mm gauge block, changing deviation from nominal central length from +80 nm to +60 nm and variation from +110 nm to +120 nm.

Regarding the principal measurand, the deviation from nominal central length, a total of 4 values out of 72 were not compatible with the final reference values per E_n, Birge ratio and chi-squared analysis, representing 5,6 % of the full set of results.

DPM did not have a relevant CMC previously, and the results originally submitted by DPM support the claimed uncertainty for gauge blocks other than 25 mm and 100 mm nominal length. However, DPM recognized after Draft A.1 that the deviation from nominal central length calculation was wrong for all 6 gauge blocks. Thus, as of Draft A.2, all deviation from nominal central length results of DPM were excluded from the SCRV calculation, and a chi-squared statistic test vs the SCRV was calculated using the corrected values submitted by DPM. The claimed uncertainty by DPM after submitting corrections is supported by chi-squared analysis.

The CMC claim is not supported for NMI (0,5 mm; 10 mm; 25 mm; 100 mm) based on outlier results in the statistical analysis, mandating corrective action to ensure compliance with CMC requirements or graying out the CMC. Overall, the largest Birge ratio of 1,248 indicates a realistic uncertainty estimation for the remaining participants.

No outliers were found in the variation measurement results of any gauge block. The Birge ratio was largest for the variation of the 100 mm gauge block at R_B 0,938, perhaps indicating gauge block length or weight induced disparities in the measurement systems of the participants compared to the Birge ratio of the shorter and lighter gauge blocks.

9 Conclusions

The following conclusions can be drawn from the EURAMET.L-S2.2.n01 supplementary comparison:

- The time taken to perform all measurements, and transport the artefacts, was 9 months longer than originally planned. The delay was partly to allow two laboratories an opportunity to measure again, rather than being excluded from the comparison, and partly due to issues related to the ATA Carnet and Customs.

- None of the gauge blocks were damaged significantly during the circulation according to the measurement results of the pilot laboratory.

- In all, the comparison has been mostly successful. A total of 5,6 % of deviation from nominal central length results are considered outliers and 0 % of variation results are considered outliers.

- The CMC claims for all participants except NMI are supported by the results of this comparison. The CMC claim for deviation from nominal central length for NMI (0,5 mm; 10 mm; 25 mm; 100 mm) is not supported. Therefore, NMI must implement corrective actions or gray out their CMC.

Appendix A – Equipment and measuring processes of the participants

The participants were asked to supply this information in a format ready for inclusion in Appendix D of the technical protocol.

METROSERT

Make and type of instrument(s)

TESA UPC 100 SN 8C02 for gauge blocks of 0,5 mm to 100 mm and TESA UPC 500 SN 9B 01 for gauge blocks of 0,5 mm to 500 mm. Resolution: 10 nm. Standard gauge blocks were calibrated by interferometry at GUM. Comparators are calibrated internally every 6 months.

The combined measurement sequence for central length and deviation is 5×(P1-P2-P1-P3-P4-P5-P6-P2). The gauge blocks were either measured by one person with both comparators on different days or by different persons on different days with undisclosed comparator and standard gauge block set in a double-blind experiment. E_n was <0,3 in every case. The results of the participant contact were used in the report.

Range of gauge block temperature during measurements & description of temperature measurement method:

Tempreature range: 19,5 °C to 20,5 °C taken as limits. The temperature of the gauge blocks was not measured for this comparison. The air temperature in the room in the vicinity of the gauge blocks was measured using an iButton DS1923 hygrothermochron with a resolution of 0,0625 °C and an expanded uncertainty of 0,3 °C.

An Almemo 470-1 surface temperature data logger with 3D printed PETG clamps was used to determine that the gauge block temperature of grade K 100 mm and 500 mm gauge blocks differed by less than the expanded uncertainty of 0,05 °C at about 19,74 °C after one full day of stabilisation when measured using the related Almemo NTC contact probes. The temperature of > 25 mm gauge blocks was observed to be close to identical to the temperature of the comparator and the air-cushioned granite table and was not significantly influenced by the oscillations in air temperature.

BEV

Make and type of instrument(s)

Mahr 626 + Millitron 1240, enhanced resolution (1 nm), dedicated software developed at BEV. Standard gauge blocks (steel, KOBA) are calibrated by interferometry in-house.

Our standard measurement sequence for the central length is: 5×(P1-P2-P2-P1), for the deviation in length: 3×(P2-P3-P4-P5-P6-P6-P5-P4-P3-P2). For this comparison each block was measured 5 to 15 times on different days by different persons.

Range of gauge block temperature during measurements & description of temperature measurement method:

Temperature range: 19,65 °C to 19.98 °C

Temperature of blocks were not measured actually, the air temperature (and humidity) in the near vicinity to the blocks is recorded and its average used to estimate the block temperature. Instrument used is a Vaisala HMT330 transmitter. The resolution of the sensor is 0,01 °C, the expanded uncertainty about 0,05 °C near 20 °C.

BIM

Make and type of instrument(s):

Gauge block comparator TESA UPD for mechanical comparison of gauge blocks with range from 0,5 mm to 100 mm. TESA UPD allows comparative measurement of two gauge blocks with same nominal length or with variation in length up to 25 mm.

Temperature device with 4 sensors type PT-100.

TESA up software program allows all length and temperature values as measured to be further processed online and makes all the necessary corrections automatically.

Range of gauge block temperature during measurements & description of temperature measurement method:

Gauge block temperature during measurements: (20 ± 0,3)°C.

Temperature device - thermometer Fluke, type 1529-R with 4 sensors, type PT-100. Two sensors are used for temperature measurement of reference and tested gauge blocks, one sensor is mounted on the measuring table and one sensor is mounted on the measuring arm close to upper probe.

FTMC

Make and type of instrument(s)

Mechanical comparator with two contacts Mahr 826.

Resolution -0,01 μm.

Range of gauge block temperature during measurements & description of temperature measurement method: the temperature of laboratory environment was measured by ALMEMO 6290-7S

Range of environmental temperature:

0,5 mm - 19,86 ÷ 20,01; 4 mm – 19,95 ÷ 20,26 mm; 10 mm – 19,88 ÷ 20,20; 25 mm -19,82 ÷ 20,02;

50 mm – 19,94 ÷ 20,12; 100 mm – 19,82 ÷ 20,09 mm.

MIRS/UM-FS/LTM

Make and type of instrument(s):

Instrument: Gauge block comparator 826, Manufacture: Mahr, Resolution: 10 nm

The traceability of the comparator was ensured by internal calibration in the laboratory (Certificate No.: 13311-LTM-23).

Range of gauge block temperature during measurements & description of temperature measurement method:

Equipment: Temperature measuring system TEMP 14, Manufacturer: JENAer Meßtechnik (Germany)

Range of gauge block temperature during measurement:

- Material sensor: (19,9262 – 20,1139) °C
- Air sensor: (19,8871 – 20,1437) °C

The temperature during the measurement was monitored on the gauge block comparator table with a material sensor, while the temperature in the measuring room was simultaneously recorded with an air sensor positioned close to device. The traceability of the measuring system was ensured by calibration in the National Temperature Laboratory - Laboratory for Metrology and Quality (LMK) of the Faculty of Electrical Engineering at the University of Ljubljana (Certificate No.: LMK0723I856-860).

NMI

Make and type of instrument(s):

TESA UPC, Model 05930015, SN 5U01

Resolution of Comparator 0,01 µm

Steel gauge blocks, grade K (TESA)

Range of gauge block temperature during measurements & description of temperature measurement method:

$T = (19{,}62 \ldots 19{,}90)\ °C, U = 0{,}05\ °C$

DPM

Make and type of instrument(s)

Gauge Block comparator EMP II, Steinmeyer, Feinmess Suhl GmbH, Germany. Range 0 – 100 mm, resolution 0.01 µm.

Digital thermometer AHLBORN, ALMEMO MA 25903S. Three channel with external probe, type NTC, resolution 0.01 °C, for gauge block temperature measurement.

Digital capacity hygrometer with thermometer TESTO Saveris H4D, with external probe, resolution 0.1 % and 0.1 °C for laboratory temperature measurement.

Range of gauge block temperature during measurements & description of temperature measurement method:

Temperature range for gauge blocks has been between 20.0 °C to 20.6 °C.

Temperature measurement method for gauge blocks is carried out by a simple contact of external probe to each gauge block surface after measurements.

LATMB

Make and type of instrument(s):

The difference measurements are carried out on a Mahr gauge blocks comparator type 826PC with a Millimar display unit, scale resolution 0,01 µm.

Range of gauge block temperature during measurements & description of temperature measurement method:

During the measurements range of temperature were 19,8 °C to 20,1 °C.

The temperature is determined directly at the gauge blocks measuring device.

The air temperature of the laboratory is continuously recorded near the measurement device.

NPL

Make and type of instrument(s)

Cary IVC-150 Gauge Block Comparator	S/N 4071
Cary IVC-153 X-Y table	(mounted on above item)
Carylabor display unit	S/N 13344

Grade K steel gauge blocks selected from two sets of master gauge blocks, S/N 100077 and 09136

Range of gauge block temperature during measurements & description of temperature measurement method:

The temperature is not recorded directly but is controlled such that the temperature during calibration will be in the range 19.8 °C to 20.2 °C. The indicator for the temperature is a self-built logging unit based on a Raspberry Pi computer, connected to digital DS18B20 temperature sensors. One sensor is near to the Cary, mounted on the steel surface plate where the gauge blocks are thermally stabilised. The logging unit and sensors are calibrated by the Temperature and Humidity group at NPL.

BFKH

Make and type of instrument(s)

- Double side tactile gauge block comparator; Cary; IVC 154; 4141; HOS-0013/2023
- Set of standard gauge blocks; Mitutoyo; BM1-122K/YJ; 2307080; 2311471
- Standard gauge block; Mitutoyo; 0,5mm; 238563; 2400977

Combined digital multi sensor thermo, and hygrometer; TITON Bt.; AEROMES 06; 201; HŐM-0065/2024 for thermometer and GAZ-0083/2024 for hygrometer

Range of gauge block temperature during measurements & description of temperature measurement method:

20 °C ± 0,5 °C;

Indirect measurement based on the temperature of the used heat compensation table, air temperature measurement with a multi sensor resistance thermometer with a resolution of 0.01 °C.

NSAI NML

Make and type of instrument(s)

TESA UPC gauge block comparator with T90 indicator

Range of gauge block temperature during measurements & description of temperature measurement method:

20.1 °C to 20.2 °C

Testo temperature system: indicator & 2 probes (1 on platten & 1 on the TESA GB comparator)

BMM

Make and type of instrument(s)

Measurement is done using the mechanical gauge block comparator, PRECIMAR 826 PC (s/n 1204/23) equipped with an electronic indication device, Mahr Millimar C 1202 (s/n: 32040184), supported with QM Soft application.

Range of gauge block temperature during measurements & description of temperature measurement method:

For the establishing declared environment temperature condition 20 ± 0,5 °C in our laboratory we are using Uniflier climate system.

Besides that, on the very measurement position we measure gauge blocks temperature using system for precise temperature measurements produced by JENAer Mestechnik GmbH, typ TEMP14 consisting of three temperatures sensor s/ns: NC269, NC270, NC271 position on the antivibrate table surface and gauge blocka dn other two on the very comparator table.

The 24 hours period of stabilization of gauge blocks is prescribed before the start of calibration.

10 References

[1] Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Oct. 14, 1999.

[2] Evaluation of measurement data — Guide to the expression of uncertainty in measurement, JCGM 100:2008. doi: 10.59161/JCGM100-2008E.

[3] K Beissner, “On a measure of consistency in comparison measurements,” Metrologia, vol. 39, no. 1, p. 59, Feb. 2002, doi: 10.1088/0026-1394/39/1/8.

[4] Figures drawn using LabPlot, https://labplot.kde.org/.