Chapter

The Data Evaluation of Interlaboratory Comparisons for Calibration Laboratories

Oleh Velychko and Tetyana Gordiyenko

Abstract

National accreditation agencies in different countries have set quite strict requirements for accreditation of testing and calibration laboratories. Interlaboratory comparisons (ILCs) are a form of experimental verification of laboratory activities to determine technical competence in a particular activity. Successful results of conducting ILCs for the laboratory are a confirmation of competence in carrying out certain types of measurements by a specific specialist on specific equipment. To obtain reliable results of ILC accredited laboratories, it is necessary to improve the methods of processing these results. These methods are based on various data processing algorithms. Therefore, it is necessary to choose the most optimal method of processing the obtained data, which would allow to obtain reliable results. In addition, it is necessary to take into account the peculiarities of the calibration laboratories (CLs) when evaluating the results of ILC. Such features are related to the need to provide calibration of measuring instruments for testing laboratories. The evaluation results for ILCs for CLs are presented. The results for all participants of ILCs were evaluated using the E_n and z indexes. The obtained results showed that for the such ILCs it is also necessary to evaluate the data using the z index also.

Keywords: interlaboratory comparison, data evaluation, referent laboratory, calibration laboratory, calibration, measurement uncertainty

1. Introduction

Participants in the International Laboratory Accreditation Cooperation (ILAC) Mutual Recognition Agreement (MRA) recognize the calibration or test results obtained by each other's accredited calibration and testing laboratories [1–4]. ILAC Policy and Procedural publications are for the operation of the ILAC MRA. ILAC has a special policy for participation in proficiency testing activities, on metrological traceability of measurement results, for measurement uncertainty in calibration [5–7]. The policy for measurement uncertainty to base on the Guide to Uncertainty in Measurement (GUM) [8–11] and retains the common understanding of the term calibration and measurement capabilities (CMCs) from the joint declaration issued by the International Bureau of Weights and Measures (BIPM) and ILAC [12]. ILAC has a special guideline for measurement uncertainty in testing [13]. This document provides guidance for the evaluation and reporting of measurement uncertainty in testing accordance with the requirements of the International Standard ISO/IEC 17025 [14].

National accreditation agencies in different countries have set quite strict requirements for accreditation of testing and calibration laboratories. Laboratory accreditation criteria in most accreditation systems include three main groups: laboratory technical equipment, personnel competence, and the effectiveness of the quality system. Interlaboratory comparisons (ILCs) are a form of experimental verification of laboratory activities to determine technical competence in a particular activity. Successful results of conducting ILCs for the laboratory are a confirmation of competence in carrying out certain types of measurements by a specific specialist on specific equipment.

To obtain reliable results of ILC accredited laboratories, it is necessary to improve the methods of processing these results. These methods are based on various data processing algorithms as required by international and regional guidelines and standards. To conduct ILC for CLs, it is necessary to take into account the relevant requirements of the international standards ISO/IEC 17025 [14] and ISO/IEC 17043 [15]. Therefore, it is necessary to choose the most optimal method of processing the obtained data, which would have a minimum number of restrictions on the application and allow to obtain reliable results. In addition, it is necessary to take into account the peculiarities of the calibration laboratories (CLs) when evaluating the results of ILC. Such features are related to the need to provide calibration of measuring instruments for testing laboratories.

ILCs for CLs are held nationally in different countries. Such ILCs are carried out to establish the competence of the CLs in calibrating various measuring instruments and working standards for various measured quantities [16–24]. For their implementation, various calibration objects are used. To evaluate the ILC data, various methods of their data processing are used [25–30], and to estimate the measurement uncertainty, the regional guidance EA-04/02 M [31] is additionally used, in addition to the ILAC documents [8, 13]. However, in addition to the method of data evaluation, it is necessary to take into account other influencing factors on the CL result of ILC. In particular, unsatisfactory ILC results for all participating CLs may be associated with a large time drift of the calibrated measuring instrument.

The growing practical need of ILCs for CLs to ensure recognition of the obtained results at both national and international levels underscores the relevance of this research.

2. The national interlaboratory comparisons for calibration laboratories

The main purpose of accredited CLs is to calibrate working standards and measuring instruments for accredited testing laboratories. Significantly more testing laboratories are accredited by national accreditation bodies than CLs. For example, at the middle of 2021, 837 testing and 35 calibration laboratories were accredited in Ukraine. This represents only 4% of accredited CLs of the total number of all accredited laboratories. Therefore, the number of ILCs for testing laboratories is objectively much larger than for CLs.

The State Enterprise "Ukrmetrteststandard" (Ukraine) as a referent laboratory (RL) organized and carry out seven ILCs for accredited CLs from 2016 to 2019 [32–35, etc]. The list of these ILCs is shown in **Table 1**. The calibration objects for these ILCs were working standards and measuring instruments for electrical quantities, and time and frequency. When carrying out comparisons, CLs calibrated objects in accordance with the requirements of the international standard ISO/IEC 17025 [14]. The total number of calibration object parameters ranged from 3 to 12. The total number of CLs with RL that took part in these comparisons ranged from 5 to 10.

ILC	Calibration object	Number of parameters	Number of participants	Period of carrying out
ILC1	Precision measuring thermocouple	AC voltage at 5 frequencies	5 labs	2016–2018
ILC2	Measures of electrical resistance (1th round)	3 nominations of resistance	8 labs	2016
ILC3	Measures of electrical resistance (2th round)	3 nominations of resistance	5 labs	2018–2019
ILC4	Precision measure of electric power	6 power factors at 2 frequencies	8 labs	2016–2018
ILC5	Low frequency signal generator	AC voltage at one frequency, total harmonic factor at 4 frequencies, 5 frequencies	4 labs	2016
ILC6	Electronic stopwatch	3 time intervals	9 labs	2016
ILC7	High-frequency signal generator	3 frequencies	10 labs	2018

Table 1.
The list of national ILCs for CLs.

In all presented ILCs, the assigned value (AV) with its uncertainty was taken as the value with its uncertainty of the RL. This was done because the RL had the best measurement capabilities among all CLs that took part in the comparisons. For many years RL has taken part in international comparisons of national measurement standards of electrical quantities within the framework of Regional Metrological Organizations (COOMET, EURAMET, and GULFMET) and had positive results. RL also had published CMCs for some electrical quantities in the BIPM Key Comparison Database [36].

A program for all ILCs was implemented in accordance with the requirements of ISO/IEC 17043 [15]. CLs that participated in the ILCs performed calibration of the measuring instruments (calibration object) provided to the RL in accordance with their own methods according to the radial scheme [4]. RL sent the calibration object to the participating laboratory, and this laboratory returned this object back to RL. In this case, the RL constantly monitored the stability of the calibration object [35, 37]. The RL determined the characteristics of the instability of the calibration object before and after its research in the CLs participating in the ILC.

In accordance with the adopted ILC programs, RL analyzed the calibration data provided by the CLs [38], in particular, analyzed the declared measurement uncertainty. The data obtained from CLs were necessarily checked by RL for their consistency. Indicators for assessing of consistency were E_n and z indexes set and defined in [4, 15]. The general algorithm of processing of the received primary data of ILCs given in [34] was used. In case of inconsistent data, RL reported this appropriate CL and analyzed the responses received from this laboratory. RL prepared a report on the comparisons, evaluating the data of all CLs. In the event that the laboratory or laboratories received inconsistent results of comparisons, RL suggested that they take the necessary corrective action.

3. The traditional data evaluation of interlaboratory comparisons

The traditional assessment of ILC data for CLs is carried out in accordance with the requirements of ISO/IEC 17043 [15]. During of the evaluation of primary data from the participating CLs, the interlaboratory deviation of the measurement results or degree of equivalences (DoE) was calculated based on the ILCs results.

The DoE for j-th CLs participant of ILC is calculated using Equation [4, 34, 38].

$$D_{labj} = x_{labj} - X_{AV}, \tag{1}$$

where x_{labj} is measured value for i-th CL; X_{AV} is AV for ILC.

Expanded uncertainty of the result of each participant $U(x_{labj})$ and expanded uncertainty of AV $U(X_{AV})$ were used to check the consistency of the primary ILC data and to calculate E_n index (E_n number) using equation

$$E_n = D_{labj} / \sqrt{U^2(x_{labj}) + U^2(X_{AV})}. \tag{2}$$

$|E_n| \leq 1.0$ indicates satisfactory performance and $|E_n| > 1.0$ indicates unsatisfactory performance [4, 15].

On **Figures 1–3** show the traditional graphical interpretation of the results of three ILCs at one of the calibration points (ILC 2–1, **Figure 1**, ILC 4–2, **Figure 2** and ILC 6–1, **Figure 3** respectively). The evaluation of primary data of all ILCs is carried out by means of the specially developed software "Interlaboratory comparisons" (Ukraine) which implements the algorithm presented in [34]. To prepare reports on ILCs, RL used specified software that allowed calculating the E_n and z indexes and constructing a graphical display of the results. The figures show the DoE with expanded uncertainty for all participating CLs in ILCs. The green dashed line shows the measurement uncertainty limits of the AV of ILC.

Only two laboratories (lab 4 and lab 6 for ILC 2–1) have an unsatisfactory result for two ILCs using the E_n index. E_n index more characterizes the reliability of measurement results of laboratories participating in the ILC, but is not always sufficient to determine the accuracy of measurement results.

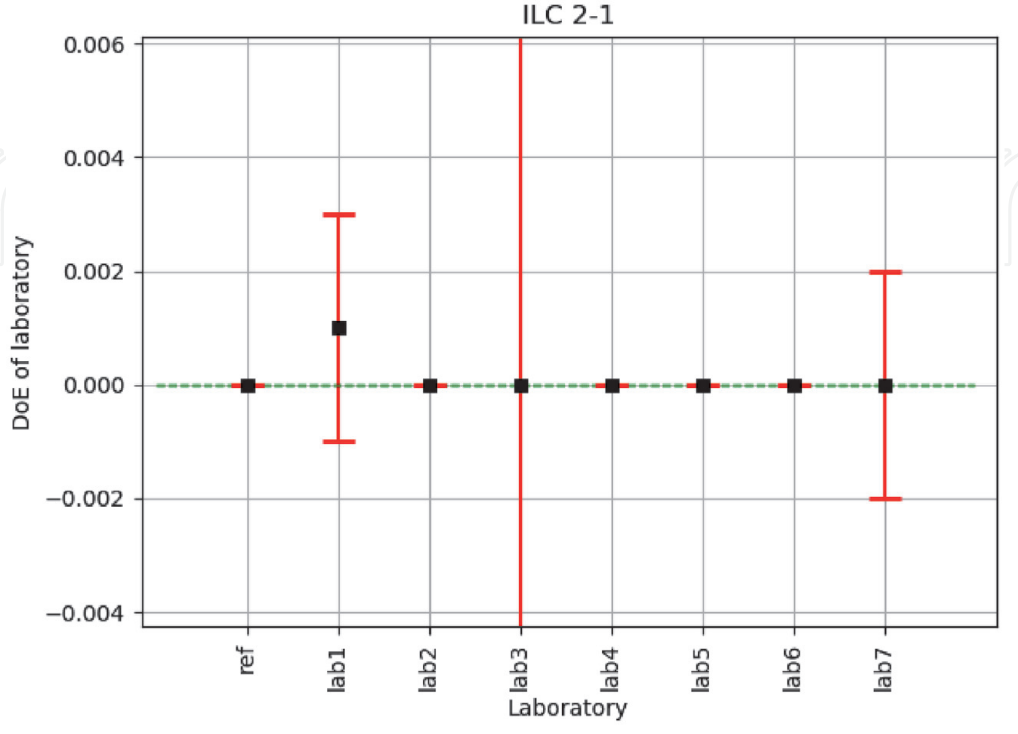


Figure 1.
DoE of CLs for ILC 2–1.

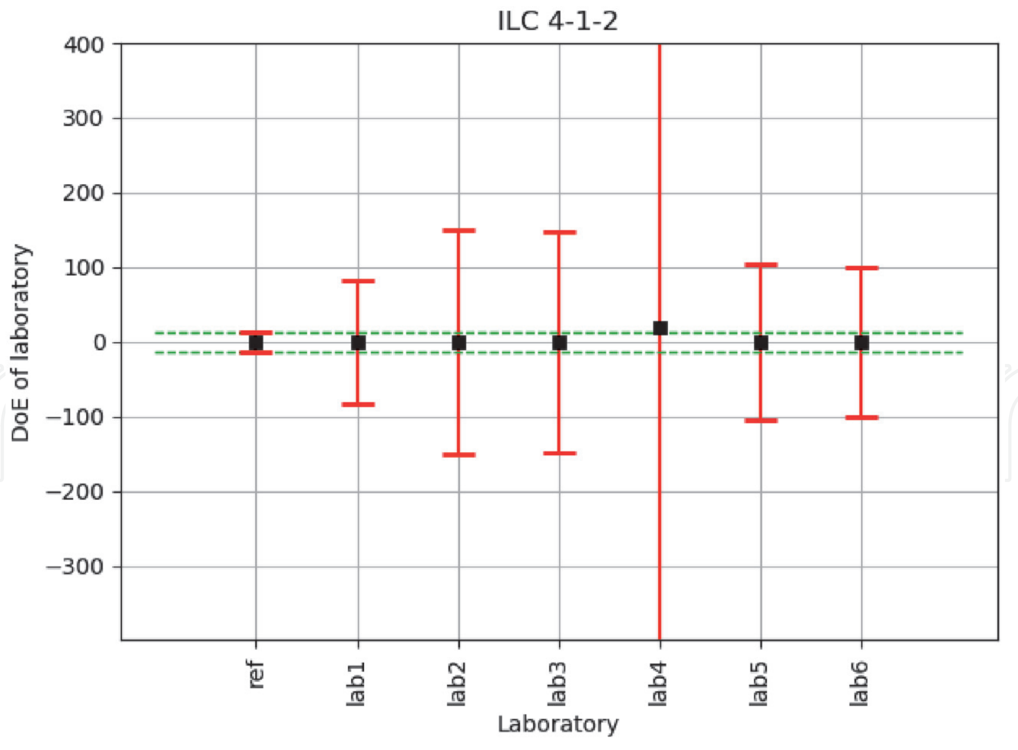


Figure 2.
DoE of CLs for ILC 4–2.

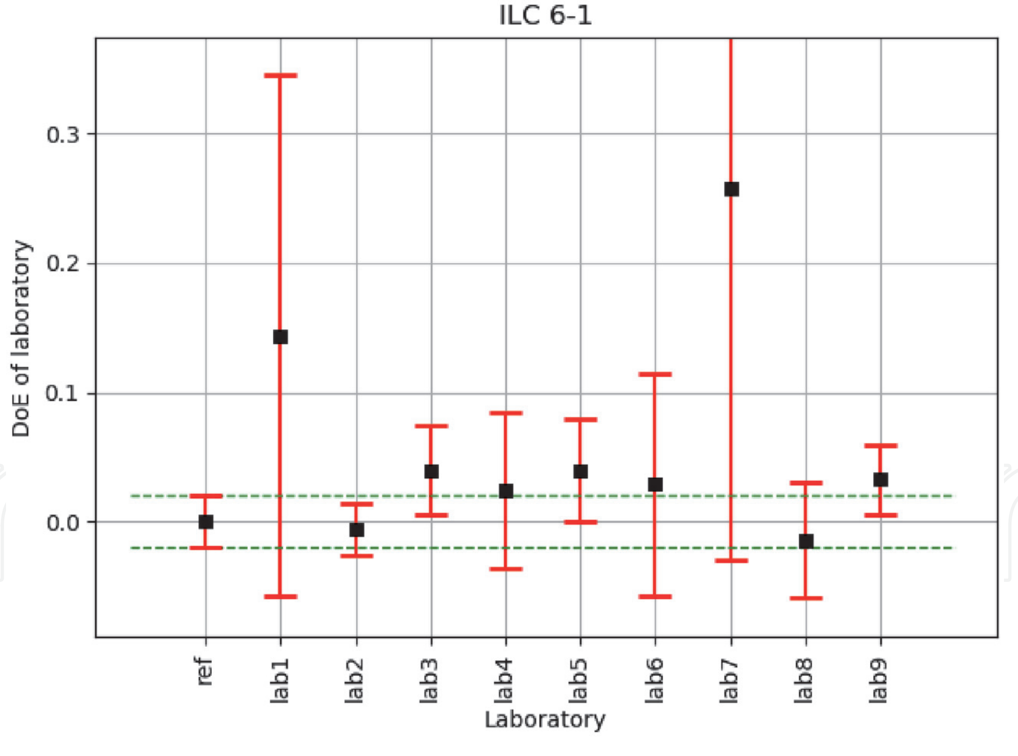


Figure 3.
DoE of CLs for ILC 6–1.

4. The additional data evaluation of interlaboratory comparisons

The consistency evaluation of data using E_n and z indicators is important not only to confirm the technical competence of laboratories participating in the ILC. This will also help to increase the accuracy of calibration by the laboratory participating in the ILC with a corresponding reduction in measurement uncertainty.

The z index compares the measurement results of all laboratories with each other and gives better information about the accuracy of measurements in laboratory. The measurement accuracy is an important characteristic for CL, therefore this index is more suitable for evaluating ILC data for CLs.

z index (z score) is calculated by the equation

$$z = D_{labj}/\sigma, \tag{3}$$

where σ is the standard deviation for qualification assessment (ILC).

$|z| \leq 2.0$ indicates a satisfactory performance characteristic and does not require adjustment or response measures, $2.0 < |z| < 3.0$ indicates a dubious performance characteristic and requires precautionary measures, and $|z| \geq 3.0$ indicates an unsatisfactory performance characteristic and requires adjustment or response measures.

In **Tables 2–8** shows the calculated results of E_n and z indexes at all points of the calibration for all ILCs. E_n and z indexes are zero for RL. Cells with unsatisfactory results are highlighted in grey in the tables. An unsatisfactory result is the excess for E_n index of the value 1, and for the z index of 2 (does not require adjustment or response measures) or 3 (requires adjustment or response measures) [4, 15].

ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4
ILC1–1	E_n	−0.854	0.444	0.438	−0.312
	z	−0.584	0.114	0.224	−2.462
ILC1–2	E_n	−0.451	0.818	1.522	−0.064
	z	−1.146	0.266	0.605	−2.090
ILC1–3	E_n	−0.645	0.882	0.987	0.022
	z	−1.167	0.483	0.781	1.895
ILC1–4	E_n	−0.147	0.238	0.129	0.452
	z	−0.025	0.019	0.018	2.503
ILC1–5	E_n	—	0.753	0.382	—
	z	—	2.265	1.939	—

Table 2.
Results of ILCs for calibration of precision measuring thermocouple.

ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
ILC2–1	E_n	0.249	−0.245	−0.008	−4.631	−0.100	2.352	−0.021
	z	2.677	−0.027	−0.996	−0.208	−0.005	0.188	−0.169
ILC2–2	E_n	−1.955	6.066	−0.018	−0.814	0.256	−0.646	−0.032
	z	−2.404	0.880	−1.460	−0.197	0.056	−0.253	−1.460
ILC2–3	E_n	−0.899	4.179	−0.086	0.171	−0.470	0.975	−0.058
	z	−2.251	1.318	−0.969	0.051	−0.127	0.367	−0.969

Table 3.
Results of ILCs for calibration of measures of electrical resistance (1-th round).

On **Figure 4** shows the graphical interpretation of the results of estimation of E_n (a) and z (b) indexes for ILC 2–1, on **Figure 5** – for ILC 4–2, and in **Figure 6** – for ILC 6–1, respectively.

ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4
ILC3–1	E_n	0.301	0.133	0.367	0.322
	z	0.117	2.658	0.152	0.961
ILC3–2	E_n	0.194	0.065	0.042	0.051
	z	0.012	2.579	0.004	0.457
ILC3–3	E_n	0.301	0.133	0.367	0.322
	z	0.117	2.658	0.152	0.961

Table 4.
Results of ILCs for calibration of measures of electrical resistance (2-th round).

ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6
ILC4–1	E_n	0.000	0.004	0.001	−0.013	0.000	0.000
	z	0.002	0.215	0.029	−2.804	0.018	0.018
ILC4–2	E_n	0.000	0.002	0.002	0.034	0.000	0.000
	z	0.001	0.046	0.046	2.875	0.004	0.006
ILC4–3	E_n	0.000	0.002	0.001	0.000	0.000	0.000
	z	0.157	2.683	0.679	−0.192	−0.105	0.157
ILC4–4	E_n	0.000	0.002	0.001	0.000	0.000	0.000
	z	0.207	2.657	1.055	−0.170	−0.075	0.207
ILC4–5	E_n	0.000	0.002	−0.001	0.119	0.000	0.000
	z	0.001	0.013	−0.009	2.859	0.000	0.001
ILC4–6	E_n	0.000	0.002	0.002	0.137	0.000	0.000
	z	0.000	0.012	0.012	2.862	0.000	0.001
ILC4–7	E_n	0.000	0.003	0.003	0.000	0.000	0.000
	z	−0.043	2.260	1.919	−0.043	0.000	−0.043
ILC4–8	E_n	0.000	0.001	0.001	0.000	0.000	0.000
	z	0.176	2.439	2.187	0.176	0.553	0.176
ILC4–9	E_n	0.001	0.002	0.000	0.000	0.000	0.001
	z	0.532	2.787	0.622	−0.009	−0.099	0.532
ILC4–10	E_n	0.000	0.002	0.000	0.000	−0.001	0.000
	z	0.108	2.570	0.569	−0.277	−0.354	0.108
ILC4–11	E_n	0.001	0.002	0.002	0.000	0.000	0.001
	z	0.331	2.462	2.068	0.095	0.253	0.331
ILC4–12	E_n	0.001	0.002	0.001	0.000	0.000	0.001
	z	0.498	2.989	1.329	0.406	0.406	0.498

Table 5.
Results of ILCs for calibration of precision measure of electric power.

ILC data	Index	Lab 1	Lab 2	Lab 3
ILC5–1	E_n	−0.235	−0.014	−0.305
	z	−2.479	−0.077	−1.044
ILC5–2	E_n	−0.050	−0.056	0.036
	z	−1.095	−1.461	1.095
ILC5–3	E_n	0.040	0.000	0.035
	z	1.206	0.000	2.412
ILC5–4	E_n	0.074	0.033	0.068
	z	1.414	1.414	2.828
ILC5–5	E_n	0.737	0.397	0.139
	z	2.399	2.181	0.727
ILC5–6	E_n	−0.087	−0.055	−0.016
	z	−2.557	−1.627	−0.465
ILC5–7	E_n	0.289	0.278	0.122
	z	2.448	2.292	1.011
ILC5–8	E_n	−0.086	0.016	−3.258
	z	−0.063	0.012	−2.326
ILC5–9	E_n	−0.284	−0.061	0.029
	z	−2.315	−0.489	0.233
ILC5–10	E_n	−0.947	−1.024	0.692
	z	−1.335	−1.442	0.975

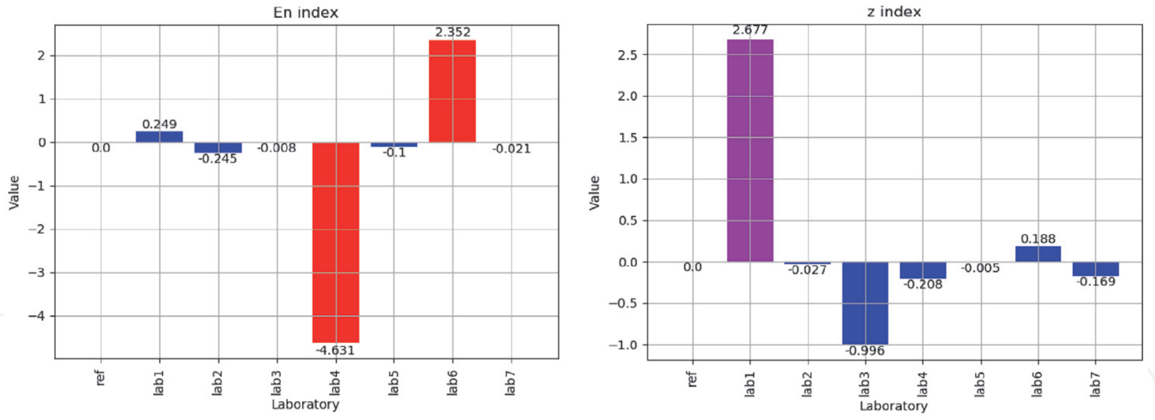
Table 6.
Results of ILCs for calibration of low frequency signal generator.

ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9
ILC6–1	E_n	0.713	−0.212	0.992	0.379	0.894	0.328	0.897	−0.290	0.982
	z	1.811	−0.075	0.503	0.302	0.503	0.365	3.245	−0.176	0.415
ILC6–2	E_n	0.998	−0.943	0.733	0.218	0.676	0.161	0.192	−0.379	0.914
	z	3.160	−0.334	0.486	0.729	0.790	0.501	1.686	−0.273	1.413
ILC6–3	E_n	0.711	−0.587	0.459	0.157	0.296	0.065	0.804	−0.273	0.754
	z	2.008	−0.193	0.468	0.908	0.468	0.289	3.012	−0.165	1.582

Table 7.
Results of ILCs for calibration of electronic stopwatch.

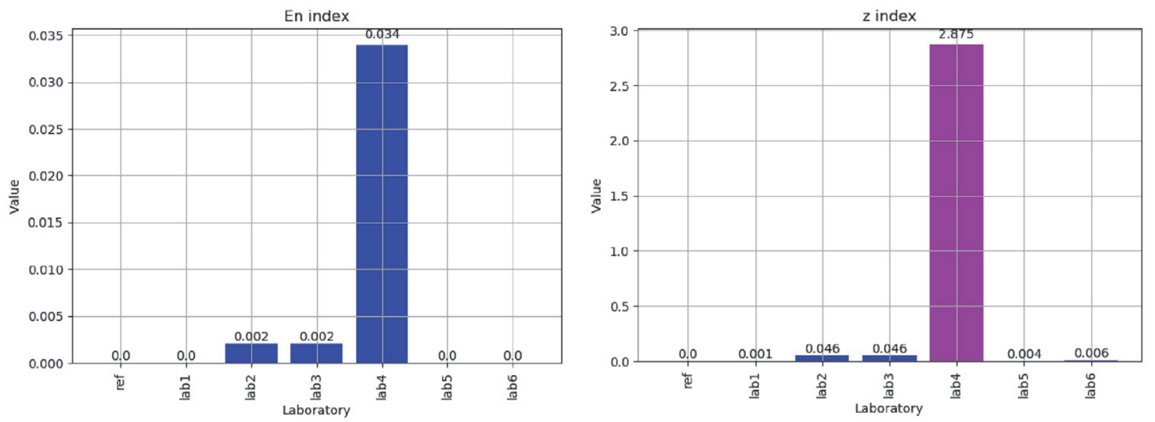
ILC data	Index	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9	Lab 10
ILC7–1	E_n	0.050	0.662	−0.072	−0.025	0.643	−0.164	−0.384	−0.039	−0.012	−0.124
	z	1.243	2.666	−0.687	−0.634	0.857	−0.124	−0.588	−0.631	−0.118	−0.433
ILC7–2	E_n	0.067	0.912	−0.076	−0.034	0.775	−0.253	−0.475	−0.068	0.005	−0.052
	z	1.261	2.533	−0.695	−0.629	0.761	−0.182	−0.725	−1.040	0.034	−0.145
ILC7–3	E_n	0.062	1.134	−0.072	−0.053	0.164	−0.160	−0.648	−0.042	−0.014	−0.104
	z	1.045	2.847	−0.746	−0.834	0.164	−0.131	−0.574	−0.427	−0.09	−0.276

Table 8.
Results of ILCs for calibration of high-frequency signal generator.



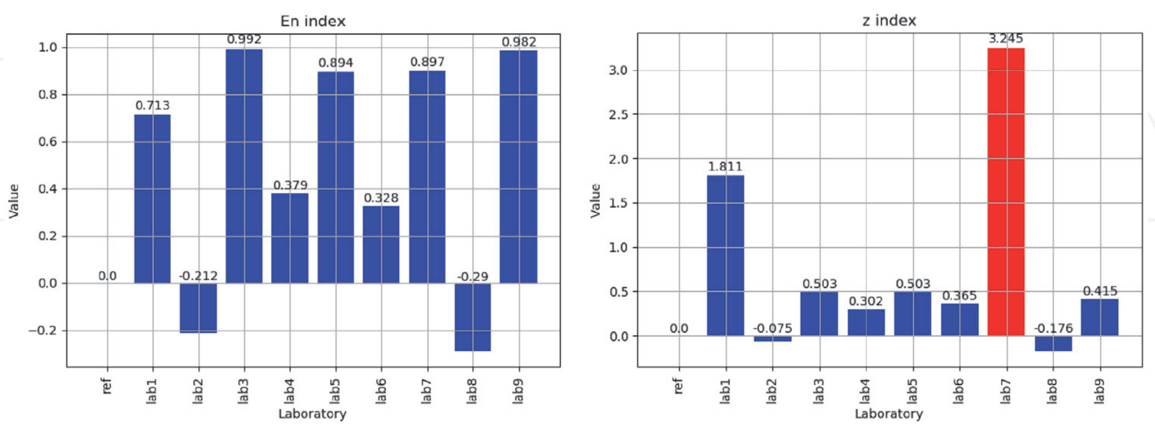
a *b*

Figure 4.
Values of E_n and z indexes for ILC 2–1: a is E_n index, b is z index.



a *b*

Figure 5.
Values of E_n and z indexes for ILC 4–2: a is E_n index, b is z index.



a *b*

Figure 6.
Values of E_n and z indexes for ILC 6–1: a is E_n index, b is z index.

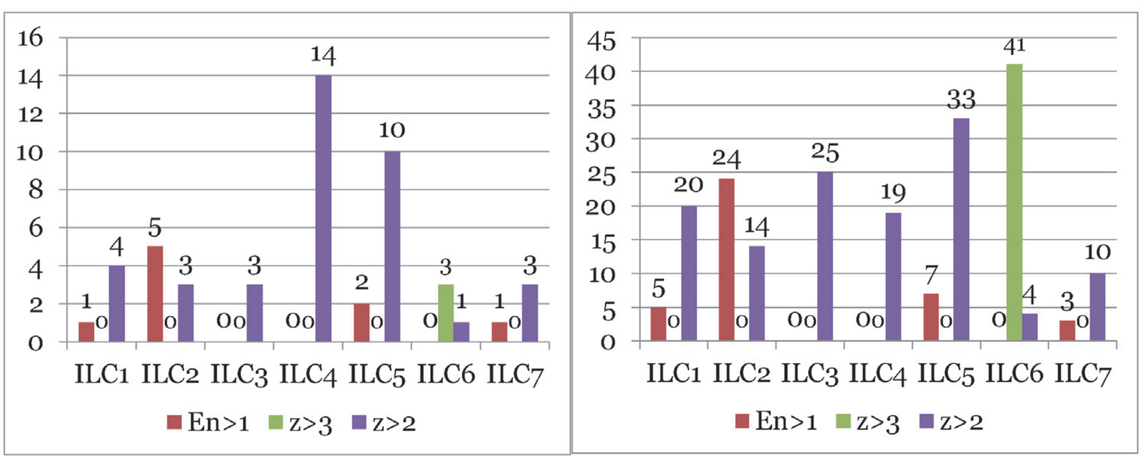
5. The summarized results of interlaboratory comparisons

The summarized results of estimation of E_n and z indexes for all ILCs are shown in **Table 9** and **Figure 7**. The percentage of discrepancies two assessments for ILCs

ILC	Number of participants*	Number of parameters	$E_n > 1$ for lab	$z > 3$ for lab	$z > 2$ for lab	Percentage of discrepancies in evaluation
ILC1	4 labs	20 points	1 point (5%)	0 point (0%)	4 points (20%)	100%
ILC2	7 labs	21 points	5 points (24%)	0 point (0%)	3 points (14%)	95%
ILC3	4 labs	12 points	0 point (0%)	0 point (0%)	3 points (25%)	100%
ILC4	6 labs	72 points	0 point (0%)	0 point (0%)	14 points (19%)	100%
ILC5	3 labs	30 points	2 points (7%)	0 point (0%)	10 points (33%)	97%
ILC6	9 labs	27 points	0 point (0%)	3 points (11%)	1 point (4%)	100%
ILC7	10 labs	30 points	1 point (3%)	0 point (0%)	3 points (10%)	97%
	Total:	**212 points**	**8 points (4%)**	**3 points (1,4%)**	**38 points (18%)**	—

*Without RL.

Table 9.
The summarized results of estimation of En and z indexes for all ILCs.



a *b*

Figure 7.
The summarized results of estimation of E_n and z indexes for all ILCs: a is absolute value, b is percentage value (%).

1, 3, 4 and 6 estimates are 100, for ILCs 2, 5 and 7 estimates are from 95 to 97. This suggests that the conclusions that can be drawn about the technical competence of the laboratories participating in these ILCs are completely inconsistent.

Only one result of ILC1 according to E_n index have inconsistency (lab 3). At the same time, 4 results of ILC1 according to z indexes have inconsistencies (lab 2 and lab 4).

5 results of ILC2 according to E_n indexes have inconsistencies (for labs 1, 2, 4 and 6). At the same time, 3 results of ILC2 according to z indexes have inconsistencies (only for lab 1).

ILC3, ILC4, and ILC6 according to E_n index have no inconsistencies. At the same time, 3 results of ILC3 according to z indexes have inconsistencies (only for lab 2), 14 results of ILC4 according to z indexes have inconsistencies (for labs 2, 3 and 4), and 4 results of the ILC6 according to z indexes have inconsistencies (for labs 2, 3, and 4), including 3 from 4 are very large ($z > 3.0$).

Only one ILC7 result according to E_n index have inconsistency (for lab 2). At the same time, 3 results of the ILC7 according to z indexes have inconsistencies (for lab 2 also).

The results of the data consistency analysis show that all ILCs, taking into account both indexes, have measurement points with unsatisfactory results. Analysis of the data taking into account the E_n index shows that only three ILCs (ILC3, ILC4, and ILC6) have satisfactory results. At the same time, analysis of the data taking into account the z index ($z > 2$) shows that all ILCs have measurement points with unsatisfactory results. ILC6 has measurement points with significantly unsatisfactory results, taking into account the z index ($z > 3$).

If we return to the analysis of **Figures 1**–**3**, it can be seen that lab 4 for ILCs 2–1 and lab 1 and lab 7 for ILC6–1 have very large declared measurement uncertainties with large DoEs. This led to unsatisfactory results, taking into account the z index. The main reason for the unsatisfactory result of lab 3 for ILCs 1–2, taking into account E_n index, is, on the contrary, a very small declared measurement uncertainty.

The general recommendation for lab 3 and lab 4 for ILC1–2, as well as for lab 4 for ILC4–2, and lab 1 and lab 7 for ILC6–1 is to revise the estimate of the measurement uncertainty, taking into account guides [8, 31]. This measurement uncertainty can be influenced by both the calibration results of the laboratory working standards and the level of competence of the laboratory personnel. Taking these recommendations into account can improve the results of that laboratories participation in other rounds of ILCs or new ILCs.

6. The influence of travelling standards instability

The travelling standards instability can affect the results of ILCs for CLs. Some works are devoted to assessing its influence, in particular compensation for its instability. The repeatability of a good measuring instrument is below 10% of its maximum error as shown in [39]. The travelling standard with 0,2% shows variations of random errors below $\tilde{x} \pm 0.02\%$ where $\tilde{x}$ is the average of the readings during calibration. This a small Type A uncertainty in relation to other components is show.

Typically, RL already takes into account the travelling standards instability in the ILC assigned value X_{AV} and its expanded uncertainty [4].

$$U(X_{AV}) = 2 \cdot \sqrt{u^2(x_{ref}) + u^2(x_{inst})}, \tag{4}$$

where $u(x_{ref})$ is the standard measurement uncertainty obtained by calibrating of travelling standard with a RL; $u(x_{inst})$ is the standard measurement uncertainty from the travelling standard instability of during ILC period

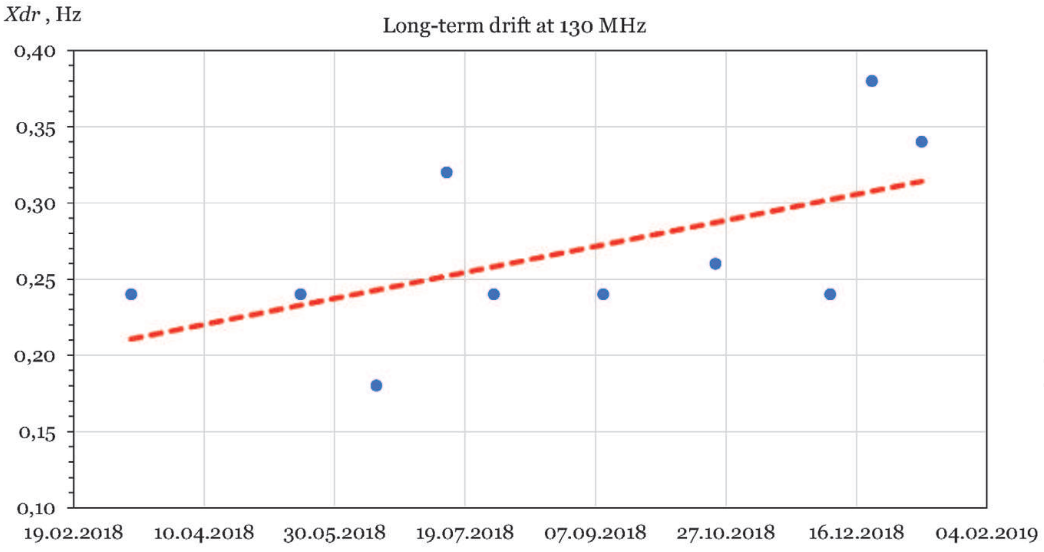
$$u(x_{inst}) = \Delta X_{\max}/\sqrt{3}, \tag{5}$$

$\Delta X_{\max}$ is the maximum change in nominal value of travelling standard during ILC period.

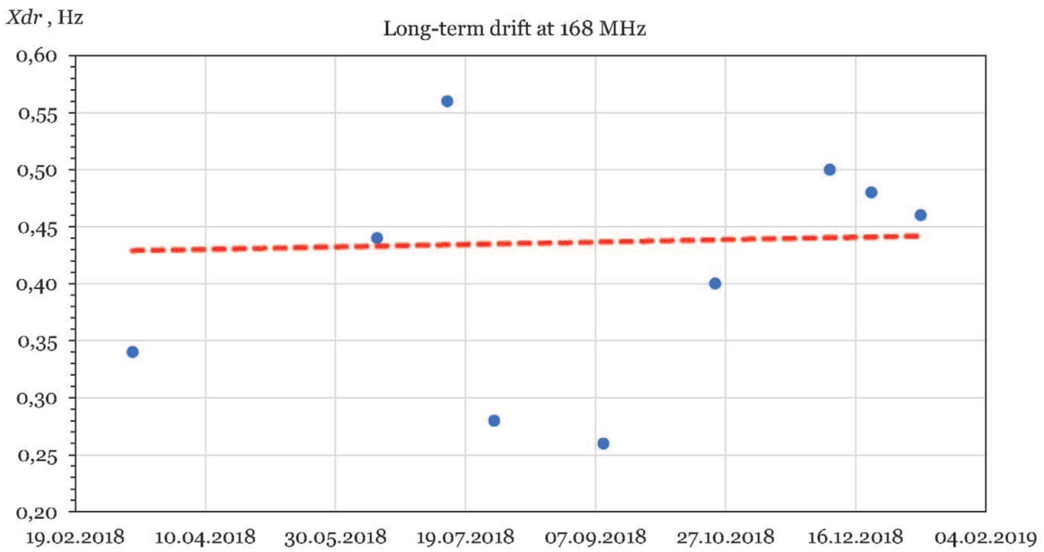
The absence of a significant effect of the travelling standards instability on the evaluation of the CL result in the ILC can be at its maximum instability, which is determined by the expression [39].

$$x_{inst} \leq \sqrt{U^2(x_{labj}) + U^2(X_{AV})}. \tag{6}$$

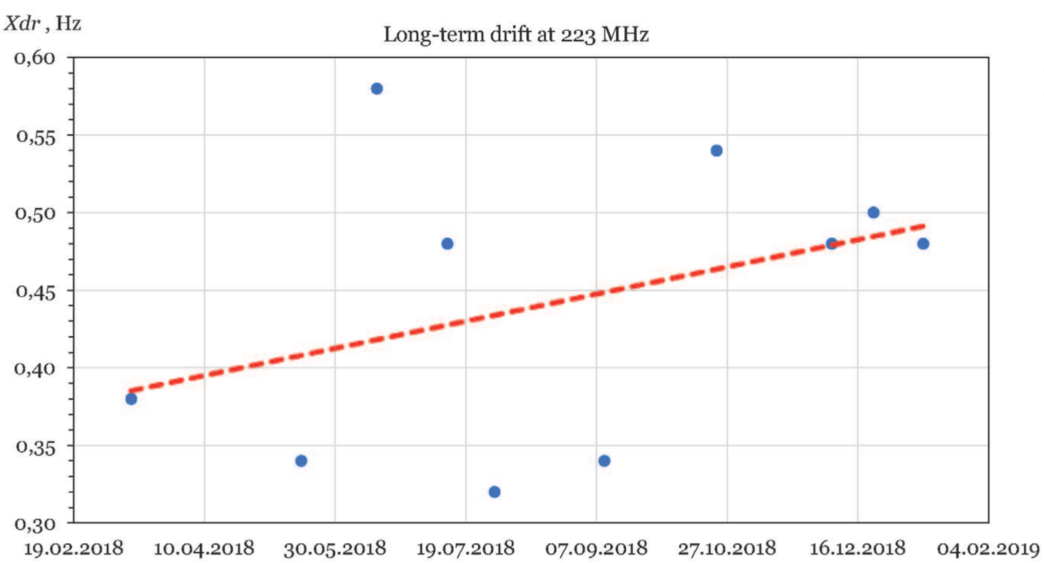
The value of the travelling standards instability can be obtained for several cases: measurements of the RL of the travelling standard in the process of carry out of ILC;



a



b



c

Figure 8.
The drift of travelling standard for ILC7: a is frequency 130 MHz, b is frequency 168 MHz, c is frequency 223 MHz.

from the technical specification for the travelling standard, measurements of the RL of the travelling standard for a long time, and etc. Results of calculating the E_n index for various options for accounting for travelling standards instability are shows in [39].

In any case, from expression (4) it follows that with an increase in the value of the measurement uncertainty associated with instability, the value of the E_n index only decreases. In this case, it can be stated that the use of a more unstable travelling standard can improve the consistency of the ILC data, which is not acceptable for CL. To carry out ILCs for CLs, it is more preferable to use working standards as calibration objects. Typically, a working standard has less instability than a measuring instrument. The use of a measuring instrument as a travelling standard can lead to somewhat distorted results of such ILCs.

An analysis RL of the travelling standard instability for all calibration points of the ILC7 is given in [33]. The drift of travelling standard for ILC7 at all frequencies is presented on **Figure 8**. The uncertainty of travelling standards instability for ILC7 is presented in **Table 10**. The contribution of the uncertainty from the long-term drift of the travelling standard to the standard uncertainty of AV for the entire duration of ILC7 is from 5.3 to 8.3% for all calibration points. Such a drift of the measuring instrument used as a calibration object is acceptable for the ILC. It does not distort the ILC results for the participating CLs.

The list of travelling standard for all ILCs and values of E_n and z indexes are shown in **Table 11.**

The use of a measuring instrument as a calibration object leads to a slight increase in the values of the z index and practically does not affect the E_n index, as can be seen from **Table 11**.

ILC7 point	Frequency (MHz)	$u(X_{AV})$ (Hz)	$u(x_{inst})$ (Hz)	Drift contribution to uncertainty AV (%)
ILC7–1	130	0.12	0.01	8.3
ILC7–2	168	0.17	0.01	5.9
ILC7–3	223	0.19	0.01	5.3

Table 10.
The uncertainty of travelling standards instability for ILC7.

ILC	Calibration object	Working standard	Measuring instrument	E_n index	z index
				Unsatisfactory (%)	
ILC1	Precision measuring thermocouple	Yes	No	5	20
ILC2	Measures of electrical resistance (1th round)	Yes	No	24	14
ILC3	Measures of electrical resistance (2th round)	Yes	No	0	25
ILC4	Precision measure of electric power	Yes	No	0	19
ILC5	Low frequency signal generator	No	Yes	7	33
ILC6	Electronic stopwatch	No	Yes	0	45
ILC7	High-frequency signal generator	No	Yes	3	13

Table 11.
The list of travelling standard for ILCs and values of E_n and z indexes.

7. The improvement of the evaluation of interlaboratory comparison results

Statistical methods for use in proficiency testing by ILCs are presented in [26, 27]. The aim of creating alternative statistics in order to improve the analysis and evaluation of ILC measurement results is research work [40]. The improvement of statistical indicators is proposed by addressing two specific issues: robustness and reliability. The proposed methodology is not traditional for ILC, but it can be used as an additional methodology for checking the results of ILC.

The following conditions are provided for data evaluation of international comparison of national standards: the travelling measurement standard is stable, the measurement results presented by laboratories are reciprocally independent, and the Gaussian distribution is assigned to a measurand in each laboratory [41–44]. The same conditions can be extended for data evaluation of ILCs for CLs. Frequently the measurement procedures for supplementary comparisons of national standards [44] are the calibration procedures of these laboratories. Such calibration procedures can also be extended to ILCs for CLs. In such a case, the calibration capabilities of the laboratory can be confirmed.

The application of z index for evaluation of CL results recommended instead of E_n index since this number is not applicable due to the difficulty in determining the AV [16]. Of course, for accredited test laboratories, it is preferable to use the services of a CL with the best calibration capabilities. Better calibration capabilities of laboratories are characterized by lower calibration uncertainties of working standards and measuring instruments. CLs with a satisfactory value of the E_n index in the ILC, but having large calibration uncertainties become uncompetitive. If we return to the analysis of **Figures 1**–**3**, it can be seen that lab 1, 3 and 7 for ILCs 1–2, lab 4 for ILC4–2, and lab 1 and lab 7 for ILC6–1 have very large declared measurement uncertainties.

The declared measurement uncertainties of CL for ILC are judged as confirmed if the following equation is satisfied [4, 43].

$$\left|D_{labj}\right| < 2u\left(D_{labj}\right). \tag{7}$$

In case the declared uncertainties CL don't confirmed during the ILC and for their confirmation it is necessary to participate in other similar ILCs.

Often, a national metrological institute or an accredited CL, which is an RL in ILC, performs high-precision calibration of working standards and measuring instruments for CLs participating in this ILC. In this case, a correlation of the obtained CL results is formed, which must be taken into account when evaluating the data of such an ILC. Covariance's are estimated by careful analysis of the uncertainty budget of CLs by the RL

$$\operatorname{cov}\left(x_{labj}, X_{AV}\right) = u_0^2, \tag{8}$$

where u_0^2 is common input to the uncertainty budgets of both results [43].

In this case, the value of the E_n index is calculated by the formula:

$$E_n = D_{labj} / \sqrt{U^2\left(x_{labj}\right) + U^2(X_{AV}) - 2\operatorname{cov}\left(x_{labj}, X_{AV}\right)}. \tag{9}$$

If the value of the E_n index meets the specified requirement (≤ 1.0), then the minimum standard measurement uncertainty, that can be claimed as calibration capability of CL participating in ILC, is:

$$u_{CC}(x_{labj}) = u_{ILC}(x_{labj}). \tag{10}$$

If the value of the E_n index not meets the specified requirement (> 1.0), then the minimum standard measurement uncertainty, that can be claimed as calibration capability of CL participating in ILC, is:

$$u_{CC}(x_{labj}) = \sqrt{\frac{D_{labj}^2}{4} + u^2(X_{AV})}, \tag{11}$$

where $u(X_{AV})$ is the standard measurement uncertainty of AV.

Correspondingly, the extended uncertainty is $U_{0.95}(x_{labj}) = 2u_{ILC}(x_{labj})$.

The same requirements can be extended for compliance (≤ 2.0) or inconsistency (> 2.0) of the value of the z index with the established requirements. In this case, the minimum standard measurement uncertainty, that can be claimed as calibration capability of CL participating in ILC, will be determined by formulas (10) and (11), respectively.

If the standard uncertainty $u(X_{AV})$ of the AV is too large in comparison with the standard deviation σ for ILC, then there is a risk that some laboratories will receive action and warning signals because of inaccuracy in the determination of the AV, not because of any cause within the laboratories. If

$$u(X_{AV}) \leq 0.3\sigma \tag{12}$$

then the uncertainty of the AV is negligible and need not be included in the interpretation of the ILC results. Further, all CLs participating in ILC shall carry out the same number of replicate measurements. This approach assumes that CLs have generally similar repeatability [26].

To evaluate the ILC data, can use z' index also [26] which calculated by the equation

$$z' = D_{labj}/\sqrt{\sigma^2 + u^2(X_{AV})} \tag{13}$$

This equation may be used when the AV is not calculated using the results reported by CLs participating in ILC. z' index shall be interpreted in the same way as z index and using the same critical values of 2.0 and 3.0.

Comparison of the equations for z and z' indexes shows that z' index for ILC will all be smaller than the corresponding z index by a constant factor of $\sigma/\sqrt{\sigma^2 + u^2(X_{AV})}$.

When the inequality established by expression (12) is satisfied, then this factor will fall in the range: $0.96 \leq \sigma/\sqrt{\sigma^2 + u^2(X_{AV})} \leq 1.00$. In this case, z' index will be nearly identical to z index, and it may be concluded that the uncertainty of the AV is negligible. When the inequality established by expression (12) is not satisfied, the difference in magnitude of the z' and z indexes may be such that some z index exceeds the critical values of 2.0 or 3.0.

8. Conclusions

To perform an ILC for CLs, RL must provide a stable working standard or measuring instrument as a calibration object and monitor its drift throughout the ILC. The use of a measuring instrument as a calibration object leads to a slight increase in the values of the z index and practically does not affect the E_n index. The application of z index for evaluation of CL results recommended instead of E_n index.

The analysis of the results of the ILC for CLs for consistency should include not only the analysis of the values of the E_n index, but also the z index. If we restrict ourselves to only the E_n index, then it is possible to get unreliable results of the ILC and not identify problems in the CL-participants of the ILC. In 3 from 7 ILCs examined, the E_n index showed completely satisfactory results, while the z index in all of these 3 ILCs revealed problematic results from the participating laboratories.

The stable travelling standard, the independent measurement results of laboratories with Gaussian distribution are main conditions for data evaluation of ILC for CLs. To participate in the ILC when declaring its measurement uncertainty, CLs must conduct a thorough analysis of the components of this uncertainty. It is necessary to take into account the correlation of the laboratory data of the participants of the ILC when evaluating its results. Covariance is estimated by carefully analyzing the CL uncertainty budget using RL.

The minimum standard measurement uncertainty that can be claimed as the calibration capability of a CL participating in an ILC can be determined in different ways depending on the value of the obtained E_n index or z index. If the standard uncertainty of AV is too large compared to the standard deviation for the ILC, there is a risk of unreliable results for some CLs.

Author details

Oleh Velychko[1]* and Tetyana Gordiyenko[2]

1 State Enterprise "Ukrmetrteststandard", Kyiv, Ukraine

2 State University of Telecommunications, Kyiv, Ukraine

*Address all correspondence to: velychko@ukrcsm.kiev.ua

References

[1] ILAC B7:10/2015. The ILAC Mutual Recognition Arrangement, ILAC, 2015; 8 p.

[2] ILAC P4:05/2019. ILAC Mutual Recognition Arrangement: Policy and Management, ILAC, 2019, 11 p.

[3] ILAC P5:05/2019 ILAC Mutual Recognition Arrangement: Scope and Obligations, ILAC; 2019, 9 p.

[4] Velychko O., Gordiyenko T. Metrological Traceability at Different Measurement Levels. Standards, Methods and Solutions of Metrology, Published by IntechOpen, London, United Kingdom, 2019; Chapter 1, pp. 1–21. DOI: 10.5772/intechopen.84853.

[5] ILAC P9:06/2014 ILAC Policy for Participation in Proficiency Testing Activities, ILAC; 2014, 8 p.

[6] ILAC P10:07/2020. ILAC Policy on Traceability of Measurement Results, ILAC; 2020, 11 p.

[7] ILAC P14:09/2020 ILAC Policy for Measurement Uncertainty in Calibration, ILAC; 2020, 14 p.

[8] JCGM 100:2008. Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), JCGM; 2008. 134 p.

[9] JCGM 100:2008. Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method, JCGM; 2008. 90 p.

[10] JCGM 104:2009. Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents, JCGM; 2009. 28 p.

[11] JCGM GUM-6:2020. Guide to the expression of uncertainty in measurement – Part 6: Developing and using measurement models, JCGM; 2020. 103 p.

[12] ILAC–BIPM Partnership [Internet]. Available from: https://ilac.org/about-ilac/partnerships/international-partners/bipm/ [Accessed: 2021-06-25].

[13] ILAC G17:01/2021. ILAC Guidelines for Measurement Uncertainty in Testing, ILAC; 2021, 12 p.

[14] ISO/IEC 17025:2017. General requirements for the competence of testing and calibration laboratories. Switzerland: ISO/IEC; 2017. 30 p.

[15] ISO/IEC 17043:2010. Conformity assessment. General requirements for proficiency testing. Switzerland: ISO/IEC; 2010. 39 p.

[16] Beckert S. F., Fischer G. E. Interlaboratory comparison of roughness measurement: Application of Algorithm A of ISO 13528:2015 in determining the designated value and the standard deviation. XXII World Congress of the International Measurement Confederation (IMEKO 2018). Journal of Physics: Conf. Series. IOP Publication. 2018;1065;082007:4. DOI: 10.1088/1742-6596/1065/8/082007.

[17] Iacobescu F., Poenaru M. M., Anghel M.-A. Reactive Power Quality Assessment through Interlaboratories Comparison. 22th IMEKO TC 4 Symposium "Supporting World development through electrical and electronic measurements", 2017; Iasi, Romania, 13–19.

[18] Poenaru M. M., Iacobescu F., Anghel M.-A. Length calibration Quality assessment through Interlaboratories Comparison. 22th IMEKO TC 4

Symposium "Supporting World development through electrical and electronic measurements", 2017; Iasi, Romania, 20–26.

[19] Poenaru M. M., Iacobescu F., Anghel M.-A. Pressure Calibration Quality Assessment through Interlaboratories Comparison. 22th IMEKO TC 4 Symposium "Supporting World development through electrical and electronic measurements", 2017; Iasi, Romania, 27–32.

[20] Poenaru M. M., Iacobescu F., Anghel A.-C., Salceanu A., Anghel M.-A. Active power quality assessment through interlaboratories comparison, 21st IMEKO TC4 International Symposium and 19th International Workshop on ADC Modelling and Testing Understanding the World through Electrical and Electronic Measurement, Budapest, Hungary, 2016, 224–228.

[21] Furuichi N., Terao Y., Ogawa S., Cordova L., Shimada T. Inter-laboratory comparison of small water flow calibration facilities with extremely low uncertainty. Measurement, 2016;91: 548–556.

[22] Bermanec L. G., Zvizdic D. Interlaboratory comparison in the pressure range from 0 to 2 MPa for accredited calibration laboratories. Int. J. Metrol. Qual. Eng., 2015;6, 307. DOI: 10.1051/ijmqe/2015021.

[23] Claudio J., Costa M. Brazilian energy interlaboratory program applicative. In: Proceedings of the XX IMEKO World Congress "Metrology for Green Growth"; 2012; Busan, Republic of Korea. IMEKO; 2012. 6 p.

[24] Briggs P. Proficiency testing for calibration laboratories. In: Proceedings of the XX IMEKO World Congress "Metrology for Green Growth"; 2012; Busan, Republic of Korea. IMEKO; 2012. 5 p.

[25] Velychko O., Gordiyenko T. The estimation of the measurement results with using statistical methods. Journal of Physics: Conf. Series. 2015; 588; 012017:6. DOI: 10.1088/1742-6596/588/1/012017.

[26] ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparisons. Switzerland: ISO; 2015. 89 p.

[27] Velychko O., Gordiyenko T. The implementation of general guides and standards on regional level in the field of metrology. Journal of Physics: Conf. Series. 2010; 238;012044:6. DOI: 10.1088/1742-6596/238/1/012044.

[28] Chunovkina A., Zviagin N., Burmistrova N. Interlaboratory comparisons. Practical approach for data evaluation. In: Proceedings of the XX IMEKO World Congress "Metrology for Green Growth"; 2012; Busan, Republic of Korea. IMEKO; 2012. 5 p.

[29] Jackson G. S., Muzikar P., Goehring B. A Bayesian approach to an interlaboratory comparison. Chemometrics and Intelligent Laboratory Systems, (2015) Vol. 141, 94–99

[30] Acko B., Brezovnik S., Sluban B. Verification of Software Applications for Evaluating Interlaboratory Comparison Results. Procedia Engineering. 2014;69, 263–272.

[31] EA-04/02 M. Evaluation of the Uncertainty of Measurement in Calibration. EA; 2013. 75 p.

[32] Velychko O., Gordiyenko T., Boriak K. Linking the Results of Inter-laboratory Comparisons for DC Electrical Resistance Measures. Ukrainian Metrological Journal. 2020;1: 4–11. DOI: 10.24027/2306-7039.1.2020.204155.

[33] Velychko O., Shevkun S., Mescheriak O., Gordiyenko T.,

Kursini S. Interlaboratory comparisons of the calibration results of signal generator. Eastern-European Journal of Enterprise Technologies. 2019;3/9(99): 14–20. DOI: 10.15587/1729-4061.2019.166504.

[34] Velychko O., Shevkun S., Gordiyenko T., Mescheriak O. Interlaboratory comparisons of the calibration results of time meters. Eastern-European Journal of Enterprise Technologies. 2018;1/9(91): 4–11. DOI: 10.15587/1729-4061.2018.121089.

[35] Velychko O., Isaiev V. Interlaboratory comparison in context of inappropriate results of voltage thermal converter calibration. Journal of Electrical Engineering and Information Technologies. 2018;3;1–2:5–12.

[36] The BIPM key comparison database (KCDB) [Internet]. Available from: http://kcdb.bipm.org/ [Accessed: 2021-06-25].

[37] Velychko O., Isaiev V. A comparative analysis of AC/DC transfer standards for comparison of national standards. Eastern-European Journal of Enterprise Technologies. Applied physics. 2018;6/5(96):14–24. DOI: 10.15587/1729-4061.2018.150459.

[38] Velychko O., Gordiyenko T. Features of the processing of results and estimation of measurement uncertainty of inter-laboratory comparison for calibration laboratories. Information Processing Systems. 2018;4(155):77–83. DOI: 10.30748/soi.2018.155.10.

[39] Sousa J. J. L., Leitão L. T. S., Costa M. M., Faria M. C. Considerations on the influence of travelling standards instability in an interlaboratory comparison program. In: Proceedings of the XX IMEKO World Congress "Metrology for Green Growth"; 2012; Busan, Republic of Korea. IMEKO; 2012. 4 p.

[40] Berni R., Carobbi C. Alternative statistical analysis of interlaboratory comparison measurement results. In: Proceedings of the XXI IMEKO World Congress "Measurement in Research and Industry"; 2015; Prague, Czech Republic. IMEKO; 2015. 5 p.

[41] Cox M. G. The evaluation of key comparison data: determining the largest consistent subset. Metrologia. 2007;44:187–200.

[42] Cox M. G. The evaluation of key comparison data. Metrologia. 2002;39: 589–595. DOI: 10.1088/0026-1394/39/6/10.

[43] COOMET R/GM/14:2016. Guidelines for data evaluation of COOMET key comparison, COOMET; 2016, 6 p.

[44] COOMET R/GM/19:2016. Guideline on COOMET supplementary comparison evaluation, COOMET; 2016, 5 p.