[1.]Floarea ŞERBANCEA, [2.]Florin NENCIU, [3.]Aurelia STĂNESCU, [1.]Cristian ŞERBANCEA

# THE INFLUENCE OF THE ASSIGNED VALUE IN THE ASSESSMENT OF TEST PERFORMANCE

[1.] IBA Bucharest, Bucuresti, ROMANIA
[2.] INMA Bucharest, Bucuresti, ROMANIA
[3.] Faculty of Management, Bucharest University of Economic Studies, Bucuresti, ROMANIA

**Abstract:** According to clause 4.4 of the ISO 17043:2010 standard, the development of the statistical project is of particular importance in the competence testing process. The criteria used in choosing the type of statistical design are clearly stated in ISO 17043 and their implementation depends on the proficiency testing provider. According to ISO 13528:2015, it is important that the statistical design is appropriate for the type and purpose of the proficiency testing scheme (PT). Entities participating in PT, according to ISO 17025:2018, must be aware of the statistical project from the stage of registration to the scheme. The aim of the study was to identify the effect produced by the choice of the assigned value in the calculation of the performance indicator of a laboratory when using the same test object. The paper presents the influence of the assigned value in the evaluation of the testing performance of laboratories according to the recommendations of the international standard ISO 13528 regarding the statistical calculation of the z–score. The study was carried out by simulating the z–score for a number of 30 participants who analyzed the ash content of wheat flour. The z–score was calculated for two assigned values (CRM value and consensus value). For the same standard deviation, significant differences were found between the z–score values.
**Keywords:** interlaboratory tests, ISO 17043, ISO 13528, z–score, assigned value

## 1. INTRODUCTION

The implementation of the test performance requirements provided in the ISO 17025 standard has led to increased confidence in the results obtained. Thus, over time, the requirements of the ISO 17025 standard were extended and targeted at activities with a risk in the quality of the results. The accreditation authority is responsible for monitoring the competence of analytical laboratories to produce correct measurement results, and accreditation obliges the laboratory to participate in relevant interlaboratory comparisons.

If initially the internal validation of the method used in the test along with the estimation of the measurement uncertainty and, respectively, the use of reference materials was sufficient, currently, the ISO 17025 standard extends quality traceability to requirements regarding competence verification by participating in interlaboratory comparisons.

Competence is expressed by performance indicators (scores), such as En, z, ζ. Therefore, without proficiency tests, compliance with the requirements of the ISO 17025 standard no longer guarantees the accuracy of the results. If it is found that the results are outside the predefined criteria in PT, through performance scores such as: z, ζ (zeta) or En number, preventive actions must be taken to comply with the accreditation requirements. The competency tests validate the statements of uncertainty and metrological traceability declared in the accreditation process. And thus, a quality proven in the accreditation process must be maintained throughout the activity. That's why the accreditation bodies ask their clients for evidence regarding the performance obtained when participating in the proficiency testing schemes. This evaluation increases the chance of continuous improvement of the measurement activity by identifying non–conformities in testing. Thus, proficiency testing is a quality assurance tool.

It is important that laboratories have information regarding the scope and availability of proficiency testing schemes in their areas of activity. This allows them to make decisions about the type of proficiency test (PT) scheme they want to participate in (Almira Softic et al., 2012). Therefore, PT schemes have multiple uses, as shown in Figure 1.

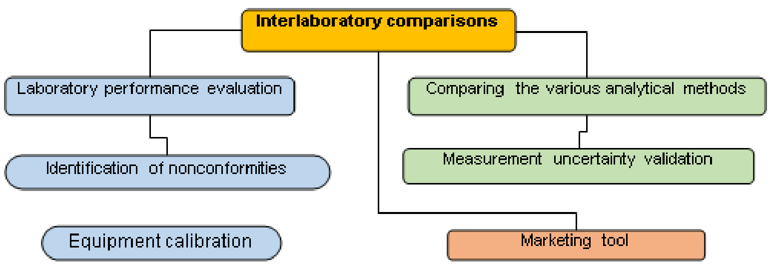


Figure 1 – Examples of PT usage according to ISO 17043


ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665

Ensuring the quality of the results is also important for the research–development activity, because the reproducibility of the results is considered a key factor in science and technology. The implementation of a quality management system based on ISO/IEC 17205 increases confidence in experimental results (Rodrigo et al., 2017).

In these conditions, the focus in the evaluation of the quality of the results moves to the field of competency tests. The activity of proficiency testing providers (PTPs) must be recognized by accrediting their activity, according to the requirements of the ISO 17043 standard. This eliminates some of the errors current methods of testing the competence of laboratories are susceptible to. In order to reduce these errors, the ISO standard places a special emphasis on the competence of the staff assigned to PT organization, data processing and interpretation of results. Thus, in the development of the statistical project, all the factors that have a major influence on the result of the competency tests are taken into account.

Heydorn demonstrates the flexibility of framing the performance of the participants in the IMEP–9 program, regarding the determination of Pb concentration in water, by comparing the scores calculated by two statistical methods, the En number and, respectively, the z score. In conclusion, Heydorn states that the z–score is a mere formality and that the actual data in ISO 13528 illustrates the extent of some errors, showing that more than half of the measurement results are considered acceptable when a robust average is used instead of a reference value (Heydorn, 2008). However, Sorbo states in his work that the values assigned by consensus were comparable to the value of the certified reference materials (CRMs) used in PT (Sorbo et al., 2020).

However, for interlaboratory comparisons (ILCs) with a limited number of participants (less than 30), when statistical methods become increasingly unreliable, it is recommended to use CRMs (Kuselman, 2010).

Also, the report presented by the Joint Research Center (JRC) on the results of ILCs organized at international level for the determination of the content of As, Cd, Pb and Hg in food supplements makes a comparison between the z and zeta scores calculated according to ISO 13528 presented in Fig. 2, for Cd (Calle, 2010). Reference values provided by NIST were used to calculate the scores and the standard deviation, for PT, was set by the Advisory Council of this ILC at 15% of the reference values for Cd, Pb and As (Baer and Calle, 2011).

| Part Nr | x1 | x2 | x3 | x4 | Ulab | k | Mean | ulab | Technique | z | zeta |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2651 | 0.0236 | 0.0225 | 0.0221 | | 0.0049 | 2 | 0.0227 | 0.0025 | ICP-MS | -0.9 | -1.5 |
| 2670 | 0.012 | 0.012 | 0.009 | | 0.004 | √3 | 0.011 | 0.002 | ICP-AES | -3.9 | -7.1 |
| 2738 | 0.018 | 0.017 | 0.016 | | 0.004 | 2 | 0.017 | 0.002 | ETAAS | -2.4 | -4.6 |
| 2751 | 0.029 | 0.025 | 0.029 | 0.025 | | | 0.027 | | Grafite Furnance A. Atómic | 0.2 | |
| 2752 | 0.0225 | 0.0226 | 0.0227 | | 0.0009 | 0.952 | 0.0226 | 0.0009 | ICP-MS | -1.0 | -3.6 |
| 2753 | 0.023 | 0.031 | 0.026 | | 0.004 | 2 | 0.027 | 0.002 | ETAAS | 0.1 | 0.1 |

Figure 2 – The z–score and score ζ –score (Baer and Calle, 2011)

Within the European IMEP–35 project coordinated by the Joint Research Center, a PT scheme was organized that supported the European Council Directive 79/768/CEE (1976). 400 laboratories from 12 countries participated in the PT scheme.

The performance of the laboratories was expressed using two scores z and, respectively, ζ (zeta). The statistical calculation differences between the two scores produce different results (Fig.3). Thus, 75% of the participants achieved a satisfactory z–score while only 50% of them were able to attain a satisfactory zeta score. (Snell et al., 2013).

The study attributed this discrepancy to the uncertainties reported by participants.

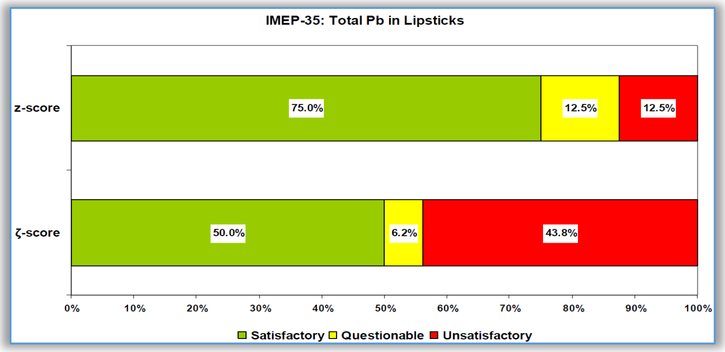


Figure 3 – The z–score and score ζ –score (Snell et al., 2013)

If the number of participants is small, the assigned value should ideally be determined independently of the participants' results. Table 1 shows a sequence of methods by some PT providers to establish assigned values, eliminate outliers and determine standard deviations (Visser, 2006).


ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665

Table 1. Assigned values, outlier tests and standard deviations used by some PT–providers (Visser, 2006)

| PT–provider | Country | Selected assigned value | Selected outlier test(s) | Selected standard deviation |
|---|---|---|---|---|
| Aquacheck [17] | UK | Mean | None | Error threshold |
| WEPAL [28] | NL | Trimmed mean | None | Actual |
| WASP [27] | UK | Trimmed mean | z>2 = excluded | Actual |
| SMPCS [26] | NL | Robust | None | Fixed target |
| RIZA [25] | NL | Mean | Cochran, Grubbs | Actual |
| KIWA [22] | NL | Mean | Grubbs, Veglia | Actual |
| FAPAS [19] | UK | Trimmed mean | ANOVA | Fixed target |
| CHEK [18] | NL | Mean | Cochran, Grubbs | Horwitz |

In their review on PTs for laboratories, the authors of the article specify that the consensus value provides useful information for researchers in the field of scientific data distribution and the development of PT schemes (Medeiros and Schwengber, 2014).

## 2. MATERIALS AND METHODS

The paper presents the influence of the value assigned in the assessment of test performance (PT) according to the recommendations of the international standard ISO 13528:2016 regarding the statistical calculation of the z score for a number of 30 participants who simultaneously analyzed the ash content of an CRM for the flour matrix of wheat.

The z–scores obtained by using preset values (CRM value) were compared with the z–score results for a consensus value of the participants' results.

### Test object

For ILC, reference material made in PTP laboratories, accredited according to ISO 17034, was used. The homogeneity and stability of the matrix were evaluated in PTP's ISO 17025 accredited laboratories (Serbancea et al., 2021).

### z–score

The competence of PT participants is expressed through performance indicators. The most used indicator is the z–score, which shows the deviation of the result from an assigned value. The z–score is a standardized measure of performance, calculated using the participant's score, assigned value, and standard deviation to assess proficiency.

According to ISO 13528, the assigned value ($x_{pt}$) and, respectively, the deviation of the result ($\sigma_{pt}$) can be established by five methods (Table 2).

Table 2. Methods for determining the assigned value ($x_{pt}$) and the standard deviation ($\sigma_{pt}$), according to ISO 13528

| | Assigned value ($x_{pt}$) | Standard deviation ($\sigma_{pt}$) |
|---|---|---|
| 1 | CRM, use expert's experience | determined using expert's experience |
| 2 | by forms | from previous rounds of a PT scheme |
| 3 | reference laboratory | result using a general model |
| 4 | expert laboratory result | reproducibility standard deviation from the previous PT |
| 5 | consensus (participant results) | from data obtained in the same round of a PT scheme |

### Interpreting the z–score value

The z–score can generate an alarm signal, depending on the values obtained. Interpretation of the z–score and actions is done as follows (Table 3).

Table 3. Interpreting the z–score value

| z–score value | Performance | Action generated |
|---|---|---|
| $\|z\|\leq 2$ | satisfactory | no alarm signal |
| $2 < \|z\| < 3$ | questionable | alarm signal, review procedures |
| $\|z\| \geq 3$ | unsatisfactory | Action signal |

### Assigned value ($x_{ref}$)

The assigned value ($x_{ref}$) plays an important role in determining the measurement performance of laboratories. Most frequently, the determination of the assigned value used in PT is done by the consensus of the participants. In the present study, two methods described in ISO 13528 were used, as follows:

### Case I – use of CRM

$$\mathbf{z} = \frac{\mathbf{x_i - x_{pt}}}{\boldsymbol{\sigma}_{\mathbf{pt}}} \quad (1)$$

where: $x_i$ – is the result reported by participant i
$x_{pt}$ – is the CRM value
$\sigma_{pt}$ – is the standard deviation (standard deviation determined by expert group)


ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665

When the assigned value is the consensus value of the participants' results, the z–score is calculated according to the equation 2. In this case, to reduce the negative effect of aberrant results on performance, it is recommended to use the robust statistical methods described in Annex D of ISO 13528.

- Case II – does not use CRM

$$z = \frac{x_i - x*}{s*} \quad (2)$$

where: $x_i$ – is the value obtained by laboratory i;
x* – is the robust average value (median);
s* – is the robust standard deviation (of the consensus value)

## 3. RESULTS

The influence of two assigned values ($x_{ref}$) for the same standard deviation ($\sigma_{pt}$) established by the proficiency test provider's expert group for a number of 30 participants was evaluated. The experimental data used in the study to calculate the z–score are presented in Table 4.

Table 4. Experimental conditions for calculating the z–score

| Version 1 | Version 2 |
|---|---|
| $z=(x_i-x_{CRM})/\sigma_{pt}$ | $z=(x_i-x_{cons})/\sigma_{pt}$ |
| $x_{CRM} = 0.63$ | $x_{cons} = 0.56$ |
| std. deviation ($\sigma_{pt}$) = 0.032 | std. deviation ($\sigma_{pt}$) = 0.032 |

The results obtained are presented in Table 6, after eliminating the outliers by checking the approximate symmetry of the distribution of results ($x_i$) presented in Table 5.

Table 5. Distribution of the results of the PT participants

| Participant code | Participant results (Xi) | Distribution of results | Participant code | Participant results (Xi) | Distribution of results |
|---|---|---|---|---|---|
| 22 | 0.41 | 0.580958 | 30 | 0.56 | 5.899173 |
| 21 | 0.45 | 1.747836 | 10 | 0.57 | 5.775207 |
| 1 | 0.47 | 2.657226 | 11 | 0.57 | 5.775207 |
| 19 | 0.47 | 2.657226 | 12 | 0.57 | 5.775207 |
| 2 | 0.48 | 3.170151 | 14 | 0.57 | 5.775207 |
| 6 | 0.48 | 3.170151 | 16 | 0.57 | 5.775207 |
| 9 | 0.48 | 3.170151 | 24 | 0.57 | 5.775207 |
| 3 | 0.54 | 5.762532 | 26 | 0.57 | 5.775207 |
| 5 | 0.54 | 5.762532 | 4 | 0.59 | 5.181975 |
| 8 | 0.54 | 5.762532 | 15 | 0.6 | 4.749487 |
| 7 | 0.55 | 5.894855 | 27 | 0.6 | 4.749487 |
| 23 | 0.55 | 5.894855 | 28 | 0.61 | 4.258497 |
| 17 | 0.56 | 5.899173 | 13 | 0.62 | 3.735291 |
| 25 | 0.56 | 5.899173 | 20 | 0.71 | 0.427131 |
| 29 | 0.56 | 5.899173 | 18 | 0.74 | 0.139602 |

The results had a normal distribution and no outliers were identified (see Figure 4).

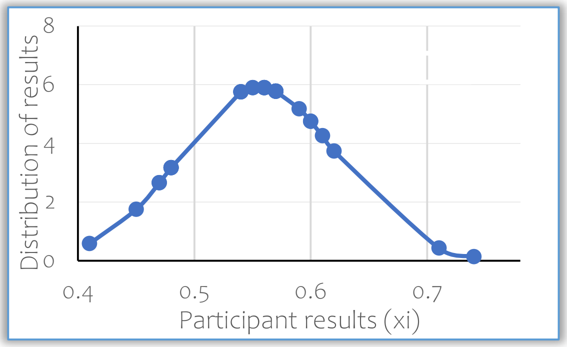


Figure 4 – Distribution of results ($z_i$)

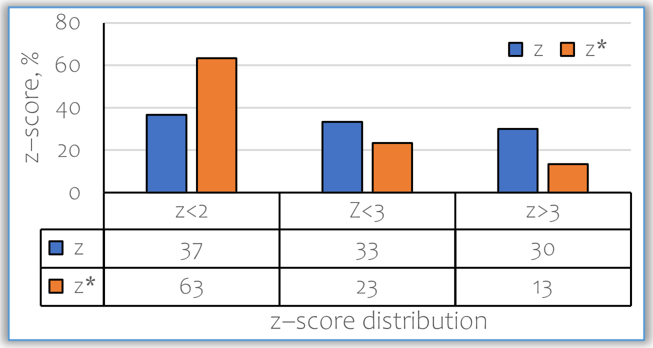


Figure 5 – The z–score distribution

From Table 6 it can be seen that the values of the z score, calculated under the experimental conditions presented in version 1 and, respectively, version 2, from Table 4, are influenced by the assigned value used. Figure 5 shows that 30% of the participants have an unsatisfactory score (z>3) when the assigned value is $X_{CRM}$, compared to only 13% when a consensus value ($X_{CONS}$) is used, calculated as the median of the results. In the graphic representations of the z–scores (Figure 6 and Figure 7) the assigned value influence can be more clearly observed. This lets us conclude that the evaluation of the participant's performance can be subjective.


ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665

Table 6. The z–score for PT participants

| Participant code | Participant results (Xi) | z– score | z* –score | Participant code | Participant results (Xi) | z–score | z*– score |
|---|---|---|---|---|---|---|---|
| 1 | 0.47 | –5.00 | –2.81 | 16 | 0.57 | –1.88 | 0.31 |
| 2 | 0.48 | –4.69 | –2.50 | 17 | 0.56 | –2.19 | 0.00 |
| 3 | 0.54 | –2.81 | –0.63 | 18 | 0.74 | 3.44 | 5.63 |
| 4 | 0.59 | –1.25 | 0.94 | 19 | 0.47 | –5.00 | –2.81 |
| 5 | 0.54 | –2.81 | –0.63 | 20 | 0.71 | 2.50 | 4.69 |
| 6 | 0.48 | –4.69 | –2.50 | 21 | 0.45 | –5.63 | –3.44 |
| 7 | 0.55 | –2.50 | –0.31 | 22 | 0.41 | –6.88 | –4.69 |
| 8 | 0.54 | –2.81 | –0.63 | 23 | 0.55 | –2.50 | –0.31 |
| 9 | 0.48 | –4.69 | –2.50 | 24 | 0.57 | –1.88 | 0.31 |
| 10 | 0.57 | –1.88 | 0.31 | 25 | 0.56 | –2.19 | 0.00 |
| 11 | 0.57 | –1.88 | 0.31 | 26 | 0.57 | –1.88 | 0.31 |
| 12 | 0.47 | –5.00 | –2.81 | 27 | 0.63 | 0.00 | 2.19 |
| 13 | 0.62 | –0.31 | 1.88 | 28 | 0.61 | –0.63 | 1.56 |
| 14 | 0.57 | –1.88 | 0.31 | 29 | 0.56 | –2.19 | 0.00 |
| 15 | 0.61 | –0.63 | 1.56 | 30 | 0.56 | –2.19 | 0.00 |

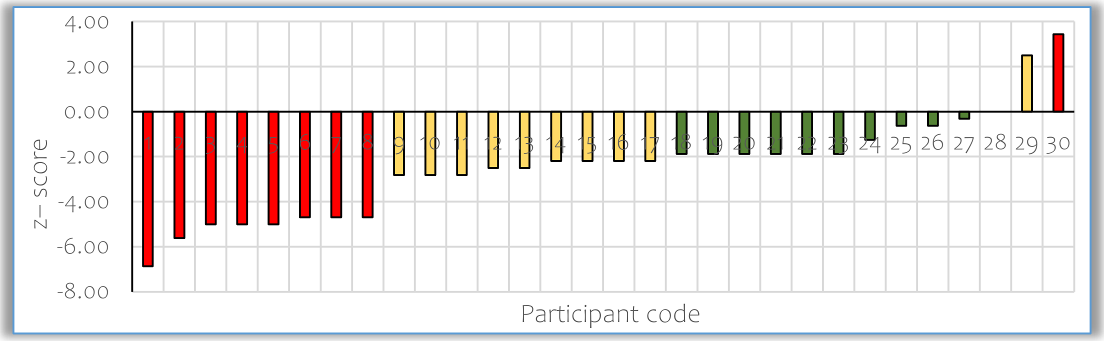


Figure 6 – The z– score

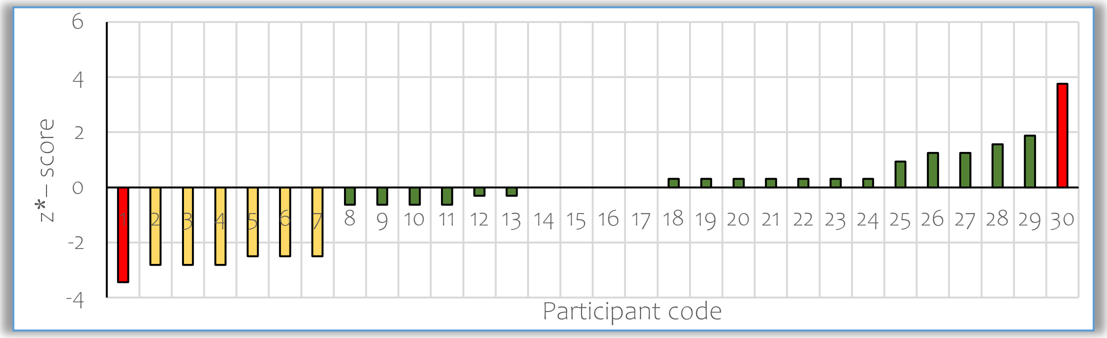


Figure 7 – The z*– score

## 4. CONCLUSIONS

According to clause 4.2 of the ISO 17043: 2017 standard, the development of the statistical project is of particular importance in PT schemes. Thus, statistical processing occupies an important place in PTP's activity and it cannot be subcontracted, the entire responsibility being assumed by the team assigned to develop the statistical project.

As shown, in certain situations, the use of consensus determined assigned value can be biased in favor of the participants of a given PT round. This can lead to situations where participants, who would otherwise achieve unsatisfactory results by using a CRM assigned value, to achieve questionable or even satisfactory results.

This would give a false representation of quality. As shown in the article, the use of CRM based assigned values eliminates this error, as long as a correctly determined standard deviation is used.

The results of the study suggest that the test performance of a participant in PT schemes is thus influenced by the assigned value used in calculating the z–score. Therefore, the performance and choices of the PT provider have a major impact on the assessment of test performance.

Also, the paper argues the importance of the statistical project and the accredited status of the PTP in the correct evaluation of the test performance.


ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665

**Acknowledgement**

This paper was supported by a grant offered by the Romanian Minister of Research, Innovation and Digitalization, as Intermediate Body for call POC/78/1/2/, project number SMIS2014 + 136213, acronym METROFOOD–RO, and Romanian Minister of Research, Innovation and Digitalization, as Core Program, contract 22N/2019, project PN 19 02 04 01.

**Note**: This paper was presented at ISB–INMA TEH' 2022 – International Symposium on Technologies and Technical Systems in Agriculture, Food Industry and Environment, organized by University "POLITEHNICA" of Bucuresti, Faculty of Biotechnical Systems Engineering, National Institute for Research–Development of Machines and Installations designed for Agriculture and Food Industry (INMA Bucuresti), National Research & Development Institute for Food Bioresources (IBA Bucuresti), University of Agronomic Sciences and Veterinary Medicine of Bucuresti (UASVMB), Research–Development Institute for Plant Protection – (ICDPP Bucuresti), Research and Development Institute for Processing and Marketing of the Horticultural Products (HORTING), Hydraulics and Pneumatics Research Institute (INOE 2000 IHP) and Romanian Agricultural Mechanical Engineers Society (SIMAR), in Bucuresti, ROMANIA, in 6–7 October, 2022

**References**

[1] Medeiros de Albano Filipe & Carla Schwengber ten Caten, Proficiency tests for laboratories: a systematic review, Accreditation and Quality Assurance volume 19, pages 245–257 (2014)
[2] Baer Ines, Håkan Emteborg, Beatriz de la Calle "Results from two interlaboratory comparisons on the measurement of trace element contents in food supplements – State of the art of control laboratories in Europe", Food Chemistry 126 (2011) 1498–1504
[3] Joint Research Centre, Institute for Reference Materials and Measurements, Calle, B., Baer, I., Sander, L., IMEP–28, total arsenic, cadmium, lead and mercury in food supplements interlaboratory comparison report, Publications Office, 2010
[4] Heydorn Kaj, 2008, The application of data from proficiency testing to laboratory accreditation according to ISO 17025, Accred Qual Assur (2008) 13:703–709
[5] Joint Research Centre, Institute for Reference Materials and Measurements, Snell, J., Emteborg, H., Tumba–Tshilumba, M., et al. (2013) IMEP–35 : determination of total lead in lipsticks : interlaboratory comparison report. Publications Office
[6] Rodrigo S. Neves, Daniel P. Da Silva, Carlos E. C. Galhardo, Erlon H. M. Ferreira, Rafael M. Trommer and Jailton C. Damasceno, 2017, Key Aspects for Implementing ISO/IEC 17025 Quality Management Systems at Materials Science Laboratories
[7] Sorbo, Angela, Ciprotti, Maria, Colabucci, Andrea, D'Amato, Marilena, Di Gregorio, Marco, Luswergh, G Fornari, Turco, Anna C. and Ciaralli, Laura. "Proficiency testing as a tool to assess quality of data: the experience of the EU Reference Laboratory for chemical elements in food of animal origin" Pure and Applied Chemistry, vol. 92, no. 3, 2020, pp. 383–390
[8] Serbancea F., Mărculescu O., Nenciu F., Stanescu A., Feasibility Study Regarding the Production of the Reference Material RM001F–IBA Wheat Flour, ISB–INMA TEH' 2021–International Symposium–Agricultural and Mechanical Engineering, pp.216–223
[9] SR EN ISO 17043: 2010 Conformity assessment — General requirements for proficiency testing, International Organization for Standardization, Geneva, Switzerland
[10] SR EN ISO 17034: 2016, General requirements for the competence of reference material producers, International Organization for Standardization, Geneva, Switzerland
[11] ISO 13528 (2015). Statistical methods for use in proficiency testing by interlaboratory comparison, International Organization for Standardization, Geneva, Switzerland
[12] ISO/IEC 17025:2018 General Requirements for the Competence of Calibration and Testing Laboratories, International Organization for Standardization, Geneva, Switzerland
[13] Kuselman Ilya, Aleš Fajgelj, 2010, IUPAC/CITAC Guide: Selection and use ofproficiency testing schemes for a limitednumber of participants—chemical analytical laboratories (IUPAC Technical Report), Pure Appl. Chem., Vol. 82, No. 5, pp. 1099–1135, 2010
[14] Visser Rob G., Interpretation of interlaboratory comparison results to evaluate laboratory proficiency, Accred Qual Assur (2006) 10: 521–526

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665