Inter-laboratory proficiency testing of the measurement of gypsum parameters with small numbers of participants

M. Kotyczka-Moranska¹ · M. Mastalerz¹ · A. Plis¹ · M. Sciazko¹

Received: 25 May 2020 / Accepted: 26 October 2020 / Published online: 12 November 2020
© The Author(s) 2020

Abstract

Conducting proficiency tests with large numbers of participants are not a problem because the ISO standard 13528:2015 describes many indicators that allow for proper assessment. However, there are proficiency testing (PT) schemes that involve only few participants. In such situations, the same difficulties exist for the selection of proper indicators and the criteria for assessing the participants' achievements. The Institute for Chemical Processing of Coal in Zabrze, Poland organised a round of PT of the determination of gypsum parameters. The results from six participants from Polish research entities and independent laboratories and methodology for the organisation this PT are presented. The performance evaluation criteria were determined using participants results due to the inability to use the metrologically valid procedure. The performance of the participants was evaluated using z' and $zeta$ scores.

Keywords Proficiency testing · Inter-laboratory tests · Gypsum · Small participant numbers · PR criteria

Introduction

The use of fossil fuels as a raw material is associated with the emission of pollutants into the environment, including sulphur dioxide, nitrogen oxides, and dust. Most of the power units of Polish power plants were built between 1970 and 1980. Poland's admission into the European Union (EU) forced Polish law to be updated to accord with EU guidelines, which resulted in tougher emission limits. To fulfil these obligations in the years from 1994 to 2013, 12 exhaust power plants were equipped with flue gas cleaning installations (IOS). Most commonly, these installations use wet flue gas desulphurisation, and a by-product of this treatment is synthetic gypsum. In 2017, the best available techniques (BAT) conclusions in the field of energy combustion were published. The emission limits set out in this document will become effective in 2021. To meet the requirements, a number of investments have been made in relation to the launches of new power units to replace depleted units. Along with the construction of blocks, new IOS installations are being built, and existing installations are being modernised.

It is estimated that, in 2020, the gypsum production capacity will increase by 1.25 tonne of Mg. The analytical laboratories of power plants that had previously focused on testing solid fuels and circulating waters were forced to establish the equipment base and research methodology for the field of limestone and gypsum testing. Most commonly, upon the purchase of IOS technology, the provider implements this research methodology in the form of procedures that are used to supervise the proper execution of the desulphurisation process. For this reason, the research methodology has been optimised in terms of the time required to obtain results, and this optimisation is associated with a decrease in the accuracy of the method. The final stage of the implementation and validation of these methods is inter-laboratory proficiency testing (PT), which later becomes an invaluable tool for the early detection of anomalies and the identification of potential problems. The present PT research is directed towards institutions and scientists who seek to confirm their competence in the field of gypsum parameter investigations. The obtained results will aid assessments of test qualities and possessed proficiencies. The current work can also be used to verify the qualifications of a laboratory to perform gypsum tests and thus allow that laboratory to gain the trust of the customers who use their services [1, 2]. The PT involved determining the mass fractions of the following oxides: Al_2O_3 , CaO , Fe_2O_3 , MgO , and SiO_2 .

✉ M. Kotyczka-Moranska
mmoranska@ichpw.pl

¹ Institute for Chemical Processing of Coal, Zamkowa str. 1,
41803 Zabrze, Poland

The preparation of sufficiently homogeneous PT items and homogeneity and stability investigations was conducted by a subcontracting laboratory. The Proficiency Testing Centre implemented a management system in accordance with the requirements of PN-EN ISO/IEC 17043:2010 [1] as confirmed by a certificate issued by the Polish Centre for Accreditation [3]. The first PTs were conducted at the Institute for Chemical Processing of Coal in Zabrze, Poland (IChPW) in 1999. Since 2008, a proficiency testing provider (PTP) has been included in the structure of the IChPW, and in 2017, the IChPW organised the first PT in the area of gypsum research. This PT represented a first in the area of gypsum in Poland. Other gypsum PTs have been conducted, for example, in the USA in 2006. This latter PT concerned the fire resistance of gypsum [4]. The German BAM Federal Institute for Materials Research and Testing (Bundesanstalt für Materialforschung und -prüfung) in collaboration with the German Metrology Institute (PTB) also organised an Africa-wide PT scheme for the testing of cement in 2014 [5]. This proficiency test addressed construction material testing laboratories in Africa. In recent years, PTs related to agricultural gypsum analysis have been conducted in South Africa. Moreover, PTs that particularly focused on examining the abilities of structural elements or systems to withstand fire have been conducted in the USA. These types of PTs are conducted by NIST (National Institute of Standards and Technology) [6].

As stated in the ISO 13528:2015 standard, for PTs with small numbers of participants, an overall value should be calculated and assigned according to a metrological procedure that is independent of the number of obtained results [7]. The most recommended procedure is a bilateral comparison in which expert judgements or criteria based on fitness for the desired purpose are applied. Another recommended procedure involves certified reference materials (CRMs) with known and stable parameters that can be investigated by the participants [8]. For PT in relation to gypsum, the use of CRMs is onerous due to cost issues. For each participant, the required sample weight is approximately 0.5 kg, thus, the fees for PT participation are too high. Additionally, the present work involved the first round of testing, so the use of parameters from a previous round would have been impossible.

When ideal conditions cannot be met, calculations of assigned values and their standard uncertainties must be performed based on the participants' results. While this approach is feasible, when the number of participants is too small, the results can be unreliable, and this fact must be accounted for. Therefore, the procedure for identifying outliers is crucial because the quantity of analysed data depends on this procedure. As previously noted, standard robust statistics are strongly recommended for outlier-contaminated populations; however, such statistics are not advisable for

very small data sets. In such situations, outlier analyses based on the mean or standard deviation may be preferable (point D.1.2 [7]). When the sample number N is small, extreme results cannot typically be identified as outliers with known statistical tests. The metrological approach for small samples makes outlier handling less important because the assigned values are not calculated by consensus, and scores are not expected to be based on the observed standard deviation. It should be noted that statistical tests for outliers are informative, but should not be used exclusively to eliminate data from a data set. Rather, their source should be examined with the intent to learn their cause. If a cause cannot be learned, the data might be analysed with and without the suspected outlier to learn of its presence in the data set is critical to the outcome. Moreover, the criterion presented in the aforementioned ISO standard regarding the uncertainty of the assigned value should also be met for small sets of participants whenever possible; i.e.:

$$u(x_{pt}) < 0.3\sigma_{pt} \quad (1)$$

where $u(x_{pt})$ = the standard uncertainty of the assigned value x_{pt} ; and σ_{pt} = the standard deviation of the proficiency assessment.

When the number of participants is below 12, this presented criterion is not met. Moreover, Algorithm A (ISO 13528:2015, Annex C, point C3) should be used for assessments involving at least 18 participants [7, 9]. In such cases, the participants' results should be evaluated with the z' -score estimator.

As presented in the previously cited standard [7], estimators such as the median can be applicable down to $p=2$. When $p=2$, the mean should be applied, and in the range $3 \leq p \leq 5$, the median has more advantages than the mean. For small groups of participants, the problem of estimating dispersion must be treated very carefully. Moreover, for small data sets, the highest available efficient dispersion estimate should be considered. The recommendations for robust estimates of dispersion in cases of small datasets are presented in Table 1.

In cases involving small groups of results, one of the most important features is the ability of the estimator to resist any bias caused by a minority group of discrepant results, i.e.,

Table 1 Recommendations for robust estimates of dispersion in cases of small datasets [7]

Number of participants	Standard deviation estimator
$p = 2$	$\frac{ x_1 - x_2 }{\sqrt{2}}$
$p = 3$	x_1, x_2 -test results
$p \geq 4$	$MADe$ (Median Absolute Deviation)
	M -estimate of the standard deviation

the resistance to a minor mode. Based on ISO 13528:2015, high resistance is characteristic of some estimators, including the $nIQR$, $MADe$, Algorithm A, and Q/Q_n [7]. Therefore, the authors decided to apply the estimators summarised in Table 2 to the small set of PT data.

However, as stated in point 8.2 of the aforementioned standard [7], the obtained standard deviation can be set as a value that is regulated by a technical expert in PT. For the presented gypsum PT data, technical experts compared the calculated σ_{pt} values with values that resulted from subjective standards and the perceptions of the experts, and they then established consensus values.

Experimental

Materials and preparation of the gypsum samples

The materials were prepared in an accredited subcontracting laboratory. The packaging and labelling of the materials for testing were performed in accordance with the instruction of the Proficiency Testing Centre (I/OBB/4.6/02/D “Sample packaging”).

Homogeneity and stability assessments

The method described in the ISO 13528:2015 standard was applied for the assessment of the homogeneity of the gypsum samples. According to this standard, the following activities were performed:

- (a) Three steps were used to ensure the correctness of the data for analysis:
 - The results for each test sample were examined in the order of measurement collection to identify the presence of any trends over time;
 - The PT item averages were examined according to production order;

- The Cochran test was applied to reject outliers.

- (b) The estimates of the within-sample standard deviation s_w and the between-sample standard deviation s_b (based on 13 measurement results) were estimated according to point B.3 of the standard ISO 13528:2015 Annex B [7].
- (c) The following criterion was examined:

$$s_b \leq \sqrt{c} \quad (2)$$

where:

$$c = F_1 \sigma_{allow}^2 + F_2 s_w^2 \quad (3)$$

$$\sigma_{allow} = 0.3 \cdot \sigma_{pt} \quad (4)$$

F_1 and F_2 are constants from statistical tables that are derived from the χ^2 and Snedecor's F distributions ($F_1 = \chi^2_{0.95}(g-1)$ where $\chi^2_{0.95}(g-1)$ is the value exceeded with probability 0.05 by a chi-squared random variable with $g-1$ degrees of freedom, and $F_2 = (F_{0.95}(g-1;g) - 1)/2$ where $F_{0.95}(g-1;g)$ is the value exceeded with probability 0.05 by a random variable with an F -distribution with $g-1$ and g degrees of freedom), c is the critical value for the homogeneity test, and s_w is the within-sample standard deviation. The samples are sufficiently homogeneous if the criterion is satisfied (significant at the 95 % confidence level). Because the PTs based on the gypsum samples were organised for the first time herein, the homogeneity assessment was performed based on a σ_{pt} value that was calculated from the repeatability and reproducibility standard deviations from a previous collaborative study in a subcontracting laboratory.

The second-most important PT sample parameter is the stability. The stability provides information about the effects that may influence sample parameters including, for example, external conditions. The assessment criterion for the stability of the gypsum samples was taken from the ISO standard 13528:2015 [7] as follows:

$$|\bar{y}_1 - \bar{y}_2| \leq 0.3\sigma_{pt} \quad (5)$$

where $\bar{y}_1$ is the average value of the parameter obtained from the homogeneity content, $\bar{y}_2$ is the average value of the parameter investigated during the stability test, and σ_{pt} is the standard deviation used in the PT.

The parameters used to investigate the sample homogeneity and stability were the mass fractions of Al_2O_3 , CaO , Fe_2O_3 , MgO , and SiO_2 . First, the material was collected and adequately conditioned. Following this preparation, 13 analytical samples, each weighing 500 g, were separated and placed in boxes. The homogeneity assessment was performed based on the results of the determinations of the

Table 2 Summary of estimators applied to the present set of gypsum proficiency testing data

Number of participants	Assigned value estimator	Standard deviation estimator
$p = 2$	Mean	$\frac{ x_1 - x_2 }{\sqrt{2}}$
$p = 3$	Median	$MADe$ (Median Absolute Deviation)
$12 \geq p \geq 4$	Median	Q_n

Al_2O_3 , CaO , Fe_2O_3 , MgO , and SiO_2 contents of the 13 samples with each item tested in duplicate. Table 3 contains raw data used for the homogeneity calculation and Table 4 contains the calculated values that were used for the homogeneity determination. As mentioned above, a very important step in a homogeneity assessment is the Cochran test for duplicate results among all tested samples, which is used to reject outliers. No outliers were detected in this investigation. Consequently, all PT test result items were included in the assessment of homogeneity. The homogeneity requirement expressed in Eq. (2) was satisfied.

The stability of the gypsum samples was investigated 7 weeks after the preparation of the materials for testing. In agreement with the technical expert, the least favourable method of storing the samples was assumed. The samples were stored on the worktop of the laboratory conducting the research. The conditions in the studio were constantly registered. The air temperature in the studio ranged from 22 °C to 26 °C, and humidity ranged from 33 to 42 %. Sample containers were opaque. Four samples were investigated. The criterion for the stability of the investigated samples is expressed in Eq. (5). During the stability testing, the aforementioned oxide contents were also examined. The standard deviations of the proficiency assessments σ_{pt} for the oxide

contents were calculated as presented above in the Introduction. Six participants investigated the Al_2O_3 , CaO , Fe_2O_3 , MgO , and SiO_2 contents; thus, the preferred standard deviation estimator was Q_n . The calculated σ_{pt} values for each of the oxide content are presented in Table 4. Due to the small number of participants, the calculated σ_{pt} values were judged by technical experts in this round of testing. The calculated values were compared with values that were determined based on the experience of the authors' laboratory and the experience of the experts (point 8.2 in [7]). Larger values were selected for further calculations. Raw data used to calculate the stability are presented in Table 5. The values of the parameters required to verify the stability of the gypsum samples are presented in Table 6. The stability criterion was satisfied.

Proficiency testing

The organisation of the PT of the gypsum analysis proceeded according to ISO/IEC 17043:2010 [1]. All instructions and relevant information for the participants were available at the IChPW web site. After registration in accordance with the rules of confidentiality, each PT participant received a

Table 3 Raw data used to calculating homogeneity

Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13
Al_2O_3 /mass fraction	0.254	0.246	0.235	0.262	0.248	0.263	0.255	0.26	0.257	0.252	0.255	0.251	0.269
	0.26	0.237	0.249	0.249	0.256	0.258	0.262	0.255	0.261	0.265	0.262	0.249	0.255
CaO /mass fraction	33.6	33.5	33.6	34.0	33.4	33.5	33.4	33.5	33.5	33.5	33.4	33.4	33.5
	33.7	33.5	33.4	33.6	33.3	33.6	33.4	33.5	33.5	33.3	33.6	33.5	33.4
Fe_2O_3 /mass fraction	0.104	0.099	0.097	0.109	0.104	0.098	0.101	0.100	0.103	0.106	0.100	0.107	0.104
	0.105	0.104	0.099	0.109	0.106	0.109	0.106	0.104	0.106	0.106	0.103	0.099	0.100
MgO /mass fraction	0.038	0.036	0.030	0.033	0.035	0.029	0.032	0.030	0.032	0.040	0.037	0.031	0.029
	0.033	0.037	0.035	0.032	0.031	0.042	0.033	0.038	0.035	0.035	0.032	0.042	0.033
SiO_2 /mass fraction	1.19	1.18	1.16	1.27	1.24	1.19	1.19	1.23	1.23	1.20	1.22	1.21	1.23
	1.23	1.17	1.19	1.23	1.17	1.20	1.23	1.19	1.20	1.23	1.22	1.21	1.21

Table 4 Summary of the parameters used in the homogeneity calculations

Parameter	Al_2O_3 /Value	CaO /Value	Fe_2O_3 /Value	MgO /Value	SiO_2 /Value
g	13	13	13	13	13
$\bar{\bar{x}}$	0.255	33.504	0.1033	0.0342	1.208
s_x	0.0068	0.1163	0.0026	0.0019	0.0194
s_w	0.0059	0.1126	0.0034	0.0044	0.0237
s_s	0.0053	0.0847	0.0011	0	0.0098
σ_{allow}	0.0081	0.9600	0.0028	0.0010	0.0360
F_1	1.506	1.506	1.506	1.506	1.506
F_2	0.469	0.469	0.469	0.469	0.469
c	0.0001	1.394	0.0002	0.0001	0.0022

g , proficiency test items; $\bar{\bar{x}}$, general average of the test results.

Table 5 Raw data used to calculate the stability

Parameter	1	2	3	4
Al ₂ O ₃ /mass fraction	0.249	0.256	0.251	0.242
	0.241	0.258	0.278	0.236
CaO/mass fraction	33.7	33.6	33.6	32.3
	34.1	33.6	33.5	32.6
Fe ₂ O ₃ /mass fraction	0.107	0.108	0.105	0.100
	0.108	0.104	0.100	0.101
MgO/mass fraction	0.037	0.032	0.037	0.038
	0.037	0.031	0.031	0.031
SiO ₂ /mass fraction	1.16	1.22	1.19	1.15
	1.16	1.18	1.24	1.12

code. All results, tables, and performance assessments in this paper are presented according to this code. Six participants took part in the gypsum PT. The determinations of the CaO, MgO, and SiO₂ contents were performed only by four participants, the determination of the Al₂O₃ content was performed by five participants, and the determination of the Fe₂O₃ content was performed by six participants. All results were determined in triplicate. All values, including the mean values and the expanded uncertainty of the measurements (for coverage factor $k=2$), were reported to the PTP. All uncertainties were calculated according to EA-04/16 [10]. All collected results were calculated and are presented on a dry-weight basis—due to the instructions given to the participants. Additional information regarding the applied techniques and instruments was also obtained. The participants performed the determinations according to specified standards and their own procedures. The participants used the test procedures described by the equipment manufacturers, other standards (i.e., the ICP-OES method according to PN-EN ISO 11885, and the X-ray fluorescence method according to PN-EN 15309), or their own procedures. Example fluorescent X-ray spectrometers used in the determinations include the S8 TIGER Bruker and the Rigaku ZSX Primus spectrometers. Additionally, for the ICP-OES method, a Leeman DRE spectrometer, an iCAP6500 DUO Thermo spectrophotometer, and a PerkinElmer Optima 800 spectrophotometer were used.

Table 6 Summary of the parameters used in the stability calculations

Parameter	Al ₂ O ₃ /value	CaO/value	Fe ₂ O ₃ /value	MgO/value	SiO ₂ /value
y_1	0.2551	33.204	0.1034	0.0342	1.208
y_2	0.2514	33.375	0.1041	0.0341	1.177
σ_{pt}	0.0270	3.200	0.0094	0.0034	0.120
$0.3\sigma_{pt}$	0.0081	0.960	0.0028	0.0010	0.036

Statistical analysis

Outliers

Before the statistical analyses, a preliminary appraisal of the results provided by the participants was performed to examine the correctness and consistency of the data. The assigned values of the measurands x_{pt} , the corresponding uncertainties $u(x_{pt})$, and the standard deviations σ_{pt} were determined based on the results obtained by the participants. These estimators were calculated according to the equations presented in Table 2. The expanded uncertainty of the assigned value $U(x_{pt})$ was calculated as follows:

$$U(x_{pt}) = k \cdot u(x_{pt}) \quad (6)$$

Where k is a coverage factor that was determined with the Student's t distribution and the appropriate corresponding degree of freedom and % confidence for each measurand ($k=2$).

As mentioned above, prior to the application of the performance statistics, the outlier analysis involving the calculation of the mean ($\bar{x}$) and standard deviation (s) necessitated by the small sample size was performed. The calculated data are presented in Table 7.

The bold numbers were above the $\bar{x} + s$ or below the $\bar{x} - s$. At the 68 % confidence level for the Fe₂O₃ content, two of six results were outside of these boundaries. Regarding the Al₂O₃ content two results of five are outliers. Regarding the MgO and SiO₂ contents, three of the four results for

Table 7 Results of the outlier determination based on the ($\bar{x} \pm s$)

No./Estimator	SiO ₂ (%)	Al ₂ O ₃ (%)	Fe ₂ O ₃ (%)	CaO (%)	MgO (%)
1	1.137	0.287	0.133	32.933	0.060
2	1.110	0.384	0.102	20.733	0.021
3	1.580	0.112	0.077	31.962	0.022
4	1.250	0.193	0.142	39.533	0.046
5	—	0.270	0.080	—	—
6	—	—	0.086	—	—
Mean	1.270	0.249	0.103	31.290	0.037
Standard deviation	0.216	0.102	0.030	7.801	0.019

each compound were within the range of the standard deviation. These findings indicated that the statistical analyses needed to be performed for four, three, and two results. ISO 13528:2015 [7] does not recommend the use of the 99 % confidence level for outlier rejection, thus, the PT organisers and the technical experts decided to use the 95 % confidence level. With the use of the 95 % confidence level ($\bar{x} \pm 2s$), no outliers were identified. Therefore, all of the results qualified for the statistical analyses.

Normality

Because no hypothesis regarding the normalities of the PT result distributions was taken into account, the compatibility of the results (as a group) should have been tested based on nonparametric statistics as stated in the IUPAC/CITAC Guide (Annex A, paragraph 4) [8]. However, due to specifics of the investigated material and the practical obstacles to testing CRMs, the compatibility of the results was not checked in this investigation.

Determining the assigned value and the standard uncertainty of the assigned value

As mentioned above, the assigned values and standard uncertainties of the assigned values were calculated according to equations presented in Table 2. After these calculations, the results were judged by technical experts, and the calculated values were compared with values that were determined according to the experiences of experts (Point 8.2 in [7]). The larger values were selected for the further calculations. The obtained and final results are presented in Table 8.

Calculation of the performance statistic

According to ISO 13528:2015 [7] and PN-EN IOS/IEC 17043:2011 [1] and accounting for the uncertainty condition (Eq. (1)), the achievements of the participants were calculated with the z' -score, which is expressed as follows:

$$z' = \frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}} \quad (7)$$

where x_i = the average value for a given parameter obtained by the participant, x_{pt} = the assigned value, σ_{pt} = the standard deviation for proficiency assessment, $u(x_{pt})$ = the uncertainty of the assigned value.

The following interpretation of the test results obtained by the participants was applied:

$|z'| \leq 2$ = satisfactory result

$2 < |z'| < 3$ = questionable result

$|z'| \geq 3$ = unsatisfactory result.

The calculated results are presented in Table 9.

ζ score calculations

ζ score can be useful for evaluating a participant's ability to produce results that are close to the assigned values. Knowledge of the uncertainty of the measurement and the results of research are very important for laboratories, their clients, and all institutions that use such results to make key decisions. The use of ζ scores allows for direct assessments of whether laboratories are able to deliver correct results [7]. According to ISO 13528:2015, the ζ score should be used to calculate x_{pt} from the participants' results to assess the correctness of the uncertainty value estimation because the ζ score encompasses the uncertainty of the assigned value and the uncertainty with which the participant performed the analysis. The ζ score is expressed in Eq. (8):

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \quad (8)$$

where x_i = the average value of a given parameter obtained by the participant, x_{pt} = the assigned value, $u(x_i)$ = the uncertainty with which the participant performed the analysis, $u(x_{pt})$ = the uncertainty of the assigned value.

The test results obtained by the participants were interpreted as follows:

$|\zeta| \leq 2$ = satisfactory result

$2 < |\zeta| < 3$ = questionable result

$|\zeta| \geq 3$ = unsatisfactory result.

The results for the participants who provided information about their uncertainties are presented in Table 10.

Table 10 Calculated ζ scores

Participant code	Mass fraction				
	SiO ₂ (%)	Al ₂ O ₃ (%)	Fe ₂ O ₃ (%)	CaO (%)	MgO (%)
7	− 0.52	1.95	0.43	0.11	− 2.57 ¹
8	−	− 4.75 ²	− 1.20	− 2.82 ¹	− 2.41 ¹
12	0.37	0.00	− 1.38	1.60	−
13	−	−	− 0.60	−	−

¹Questionable result

²Unsatisfactory result

Figure 1 presents the mean values of the investigated measurands of the gypsum x_i , the expanded measurement uncertainties $U(x_i)$, the assigned values x_{pt} , their expanded

uncertainties $U(x_{pt})$, and the PT standard deviations σ_{pt} . The solid red lines represent the assigned values x_{pt} , the solid yellow lines delimit the assigned intervals $x_{pt} \pm U(x_{pt})$

Fig. 1 Measurement results and uncertainties (vertical bars) of the determined measurands. **a** SiO_2 , **b** Al_2O_3 , **c** Fe_2O_3 , **d** CaO , **e** MgO . The solid lines represents the assigned values x_{pt} , the dashed lines

delimit the assigned intervals $x_{pt} \pm U(x_{pt})$ ($k = 2$) and the dotted lines delimit the target intervals $x_{pt} \pm 3 \cdot \sigma_{pt}$

Table 8 Summary of the calculated values and final results

Mass fraction	Assigned value	Uncertainty of the assigned value	Expanded uncertainty of the assigned value ¹	Calculated standard deviation for proficiency assessment	Final standard deviation for proficiency assessment
	x_{pt}	$u(x_{pt})$	$U(x_{pt})$	σ_{pt}	σ_{pt}
SiO ₂ (%)	1.193	0.149	0.299	0.0304 (Q_n)	0.239
Al ₂ O ₃ (%)	0.270	0.030	0.060	0.0313 (Q_n)	0.054
Fe ₂ O ₃ (%)	0.094	0.010	0.019	0.0118 (Q_n)	0.019
CaO (%)	32.448	4.056	8.112	1.108 (Q_n)	6.490
MgO (%)	0.034	0.004	0.009	0.0015 (Q_n)	0.007

¹ $U(x_{pt})$ was calculated with the coverage factor $k=2$

Table 9 Calculated z' values

Participant code	Mass fraction				
	SiO ₂ (%)	Al ₂ O ₃ (%)	Fe ₂ O ₃ (%)	CaO (%)	MgO (%)
1	− 0.20	0.27	1.84	−	3.14 ²
7	− 0.29	1.84	0.36	0.06	− 1.57
8	−	− 2.55 ¹	− 0.78	− 1.53	− 1.41
10	1.37	− 1.25	2.23 ¹	− 0.06	1.39
12	0.20	0.00	− 0.66	0.93	−
13	−	−	− 0.38	−	−

¹ Questionable result

² Unsatisfactory result

($k=2$), and the solid blue lines delimit the target interval ($x_{pt} \pm 3\sigma_{pt}$), which is relevant for z' score evaluation. Based on Fig. 1, the qualitative measurands and estimated measurement uncertainties were evaluated. If the measurands were determined correctly, their values should remain in the region defined by $x_{pt} \pm 2\sigma_{pt}$. In the present work, the participants achieved satisfactory z' scores.

Discussion

The qualitative measurands and estimated measurement uncertainties were evaluated based on Fig. 1. If the measurands are determined correctly, their values should remain in the region defined by $x_{pt} \pm 3 \cdot \sigma_{pt}$. In the present work, the participants achieved satisfactory z' -scores. Additionally, in accordance with ISO 13528 [7], the qualitative parameters and estimated measurement uncertainties were properly determined if their values covered the area defined by $x_{pt} \pm U(x_{pt})$. As observable in the figure, not all of the participants performed valid calculations of the expanded uncertainties of their results. Moreover, in some cases, the uncertainties were correct, but the values were

too high (for example, the code 8 participant in the case of the CaO content). Some of the participants estimated the measurement means correctly, but the uncertainties of the investigated parameters were too high or too low (for example, the code 7 participant in the case of the Fe₂O₃ content).

The calculated z' and ζ values are presented in Tables 8 and 9. Regarding the MgO content, one unsatisfactory result was obtained (participant code 1). Regarding the Al₂O₃ and Fe₂O₃ contents, two participants (codes 8 and 10, respectively) obtained questionable results. These findings indicate that the participants with codes 1, 8, and 10 should be advised to check their measurement procedures based on these warning signals.

Analysis of the ζ parameter revealed that one participant (code 8) obtained an unsatisfactory result for the Al₂O₃ determination. Three questionable values were also detected (for the CaO and MgO contents obtained from participant code 8 and the MgO content obtained from participant code 7). The use of ζ scores allows participants to identify problems with measurement uncertainties and provides information about the abilities of laboratories to deliver correct results.

Conclusions

Currently, the ability of a laboratory to conduct investigations with a high level of quality is among the most desirable skills from the customer's perspective. The opinion of the laboratory often plays a decisive role, and therefore, reliable results are the most sought-after product. Thus, it is often necessary to conduct PT among small groups of specialised laboratories, which, for accreditation purposes, are required to participate in such testing. The need to implement PT among small numbers of participants is necessarily associated with the acceptance of certain restrictions. The first such restriction is the need to adjust the scope of the research according to the costs of organising the PT. An excessively wide range of tests

may result in a drastic increase in the costs of preparing large samples and conducting homogeneity tests for a wide range of parameters. Another limitation is the need for very careful analyses of the produced data. In situations with small numbers of participants, the rejection of outliers may result in drastic changes in the assigned values. Therefore, the analysis of the results in such situations requires the very close cooperation of statisticians and technical experts in the field. Additionally, for topics of investigation for which prior PT has not been performed, the use of CRMs is not possible, and cooperation with technical experts is the only option for conducting PT. Moreover, when it is not possible to compare results across laboratories due to the highly specialised nature of a laboratory, PT involving the close cooperation of experts is the only option for confirming laboratory skills.

In the present analysis, six participants investigated the oxide contents of gypsum. Because the different oxides were examined by a small number of participants, the standard uncertainties were calculated for the assigned values and judged by technical experts. Based on the analyses of the obtained results, the experts decided to increase the standard uncertainties of the assigned values. The obvious conclusion is that all of the participants achieved satisfactory results. However, the measurement procedures of one participant require examination. A second investigation examined the ζ scores. The ζ score is strictly associated with the participant's uncertainty. Unfortunately, only four participants reported the information required for ζ score analysis. Despite this limitation, it was possible to compare the participants' uncertainties, and it can be concluded that the participants, especially the participant with code 8, should analyse their sources of uncertainty.

Acknowledgements This work was sponsored within a Statutory R&D Project of IChPW (No. 11.19.022.002.001).

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Reference

1. ISO/IEC 17043:2010 General requirements for proficiency testing. ISO, Geneva
2. EURACHEM/CITAC, C. of E.-P. W. Group (2011) Selection, use and interpretation of proficiency testing (PT) schemes, p 52
3. Polish Centre for Accreditation, certificate available at
4. Grosshandler WL (2007) Fire resistance proficiency testing of gypsum/steel-stud wall assemblies. Tech. Memo. PWRI 4075, Wind Seism. Eff. US/Japan Nat. Resour. Dev. Progr. (UJNR), Jt. Meet. 39th, 2007
5. Africa wide Proficiency Testing Scheme
6. Grosshandler WL, Manzello SL, Maranghides A, Mizukami T (2007) Analysis of inter-laboratory testing of non-loadbearing gypsum/steel-stud wall assemblies, p 40
7. ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, Geneva
8. Kuselman I, Fajgelj A (2010) IUPAC/CITAC guide: selection and use of proficiency testing schemes for a limited number of participants-chemical analytical laboratories (IUPAC technical report). Pure Appl Chem 82(5):1099–1135
9. Nyrk K (2006) Ryzyko związane z małą liczbą uczestników badania biegłości. No 6, pp 42–44
10. EA, EA-4/16 G:2003 EA guidelines on the expression of uncertainty in quantitative testing. No December, p 28

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.