PRACTITIONER'S REPORT

Liat Kimhi
Chaim Zlotnikov
Ilya Kuselman

# Local comparability of proficiency testing results: determination of concrete slump and compressive strength values

L. Kimhi
Israel Laboratory Accreditation
Authority (ISRAC),
2 Habonim St., Beit Habonim, Ramat
Gan 52522, Israel

C. Zlotnikov
ISOTOP Ltd., Ashdod,
101 Ha'adarim St., P.O. Box 2, Gedera
70700, Israel

I. Kuselman (✉)
The National Physical Laboratory of
Israel (INPL),
Danciger "A" Bldg, Givat Ram,
Jerusalem 91904, Israel
e-mail: ilya.kuselman@moital.gov.il
Tel.: +972-2-5661856
Fax: +972-2-6520797

**Abstract** A proficiency testing (PT) scheme is developed for comparability assessment of results of concrete slump and compressive strength determination. The scheme is based on preparing of a test portion/sample of a concrete in-house reference material (IHRM) at a reference laboratory (RL) in the same conditions for every PT participant. Therefore, in this scheme IHRM instability is not relevant as a source of measurement/test uncertainty, while intra- and between-samples inhomogeneity parameters are evaluated using the results of RL testing of the samples taken at the beginning, the middle and the end of the PT experiment. The IHRM assigned slump and compressive strength values are calculated as averaged RL results. Their uncertainties include the measurement/test uncertainty components and the components arising from the material inhomogeneity. The test results of 25 PT participants were compared with the IHRM assigned values taking into account both the uncertainties of the assigned values and the measurement/test uncertainties of the participants. Since traceability of the IHRM assigned values to the international measurement standards and SI units cannot be stated, local comparability of the results is assessed. It is shown, that comparability of the slump and compressive strength determination results is satisfactory, while uncertainty evaluation for slump results requires additional efforts.

**Keywords** Comparability · Proficiency testing · Concrete · Slump · Compressive strength

## Introduction

Comparability of measurement/analytical results is the realization of their property "traceability" [1], and a conclusion on the equivalence of the results ("tested once, accepted everywhere" [2]) can be made only when the results are comparable. Distributions of analytical results and hypotheses necessary for development of their comparability criteria, as well as comparability of the results obtained in proficiency testing (PT) based on the metrological approach are discussed recently in ref. [3–5]. The approach implies the use of samples of a reference material (RM) with traceable property values as test items sent to the laboratories participating in PT. Such RM is a working measurement standard to which PT results are traceable. Assuming normal distribution, comparability (equivalence of the PT results) is assessed by the bias of the mean of PT results $C_{PT}$ from the assigned/certified value $C_{cert}$ of the RM property, taking into account standard uncertainty $\sigma_{cert}$ of this value and standard deviation $\sigma_{PT}$ of the PT results. However, when traceability of the RM property value $C_{cert}$ prepared for PT is questionable, comparability of the PT results cannot be assessed in the meaning "tested once, accepted everywhere". In such cases, especially when number N of the participants is limited, a local comparability, i.e. among the participants only, is assessed [3].

The main problem of RM developing for fresh concrete testing is its inhomogeneity and instability. To overcome this problem, every participant in the CCRL Concrete Proficiency Sample Program, overseen by ASTM [6], is provided with a sample consisting of the concrete RM components. The participant mixes these components at its

laboratory, i.e. prepares the concrete (sample) independently. It allows to start the test immediately after preparing the concrete and to eliminate material instability as a source of uncertainty of the test results. However, the situation with RM inhomogeneity is more complicated. In spite of homogenizing the sample components at the Cement and Concrete Reference Laboratory (CCRL at NIST, USA) and supplying the mixing instructions to the program participants, inhomogeneity of the concrete prepared in different laboratories influences the test results distribution. Deviations of the participant results from their average/consensus value $C_{PT/av} \equiv C_{cert}$ in comparison to the standard deviation $S_{PT}$ (based on Youden plot analysis [7]) are used for evaluation of the participant proficiency. Since in general participants in the program are situated far both each from other and from CCRL (in different states and even on different continents) and their number N is big enough ($N > 100$), the program is probably optimal and can indicate broad comparability if uncertainties and traceability of the test results and of the RM properties are stated.

The purpose of the present publication is to develop a PT scheme for comparability assessment of results of concrete testing in a limited region, like Israel. Such a scheme is intended for the local laboratory accreditation body (IS-RAC in Israel) to control the performance of not numerous accredited laboratories ($N < 30$) that test concretes for the local building industry. Slump and compressive strength are chosen in the scheme as the test parameters of fresh and hardened concrete, practically the most required by the customers.

## Experimental

### Test/measurement methods

For slump determination by standard [8] a slump-cone (of tin plate, upper and bottom diameters are of 100 and 200 mm, respectively, height is of 300 mm) is filled with the concrete to be tested and is compacted completely to the upper ring of the cone manually by means of a steel rod. Then the slump-cone is being carefully lifted, also manually, during 5–10 s. The height of the remaining concrete cone is measured with a ruler and the slump determination result is calculated as the difference in the heights of the slump- and concrete cones, mm.

The compressive strength is measured by standard [9] as a pressure, MPa, applied by a special testing machine to 100 mm hardened concrete test cube in order to destroy it. To prepare the test cubes the standard requires to fill the corresponding steel forms with the concrete by hand using a steel rod or by means of a vibrating table. Afterwards, the test cubes should be stored 7 days under controlled conditions (air temperature of $21 \pm 3\ ^{\circ}$C and humidity of more than 95%) and then 21 days under standard laboratory conditions for hardening. On the 28th day the test cubes should be destroyed. In many cases customers are also

**Table 1** Composition of the concrete IHRM produced for the PT (calculated for a sample of 35 L)

| No | Component | Mass, kg |
|---|---|---|
| 1 | Crushed aggregates till 22 mm | 19.6 |
| 2 | Crushed aggregates till 14 mm | 10.85 |
| 3 | Crushed aggregates till 7 mm | 11.90 |
| 4 | Sea sand | 23.8 |
| 5 | Cement 42.5 CEM II/ALL | 10.85 |
| 6 | Water | 6.82 |

interested in compressive strength determination on the 7th day of the test cube hardening.

### Design of experiment

The Research Unit of the Department of Building Units and Materials at ISOTOP Ltd. served as a Reference Laboratory (RL) in the experiment overseen by INPL (the National Metrology Institute).

Composition of the in-house reference material (IHRM) [10, 11] developed as a local working measurement standard for the PT and corresponding to a fresh concrete of type B30 by standard [12] is shown in Table 1. The aggregates were thoroughly washed with water before the experiment, dried till constant weight at $105 \pm 5\ ^{\circ}$C, sieved and homogenized. The sea sand was also dried till constant weight at $105 \pm 5\ ^{\circ}$C, sieved (the fraction smaller than 0.65 mm was used) and homogenized. The components were stored in RL at air humidity of 45–60%.

The concrete for every PT participant (IHRM sample of 35 L) was produced by RL using the same Pan Mixer of 55 L, company "Controls", Italy, in the same conditions. Every participant had a possibility to start testing its sample from the moment when the concrete preparation was finished, like in the CCRL Program. However, the material homogeneity in this scheme is probably higher than in CCRL Program (since all the samples are prepared identically) and can be evaluated. Therefore, the concrete stability was not taken into account, while inhomogeneity parameters were studied during the experiment and included in the uncertainty budget of the assigned values of the IHRM properties.

Twenty five participants took part in the experiment ($N = 25$). Twenty nine samples were prepared by RL during two weeks in September 2005 before a season of rains in Israel influences the air humidity. RL tested the 1st, 12th, 23rd and the 29th (last) samples for the material inhomogeneity study and characterization. Other samples were tested by the PT participants according to the schedule preliminary prepared and announced.

The slump duplicate determinations were performed at RL by representatives of every participant using their own facilities and standard operating procedures (SOP) corresponding to standard [8]. Immediately after the slump determination, twelve 100 mm test cubes for compressive strength determinations were prepared by representatives

of every participant also using their own facilities and SOP corresponding to standard [9]. On the next day after preparing the hardened cubes were transferred from RL to the laboratory of the participant, where compressive strength determinations were performed both on the 7th day and on the 28th day after the sample preparation (every one of 6 replicates).

The data obtained were sent by the participants to ISRAC and re-sent from ISRAC to INPL in an anonymous form. For this reason, the participants are named in the paper by numbers $j = 1, 2, \ldots, N$ (25).

The results of the compressive strength determinations on the 7th day and some additional information (the mass of the cubes, correlation between the mass and the strength results on the 7th and on the 28th days, etc.) are not discussed further for short.

## Results and discussion

### Slump determination

#### *Homogeneity estimation, assigned value and its uncertainty*

The RL results of slump determination are presented in Table 2, where rows represent samples; $X_1$ and $X_2$ are the duplicate values; $R = |X_1 - X_2|$ is the range; $L = 2.77 u_{mRL}$ is the limit of the range at the level of confidence of 95%, $u_{mRL} = 3$ mm is the standard measurement/test uncertainty declared by RL; and $X_{avg} = (X_1 + X_2)/2$ is the average result of the slump determinations for a sample. There is also a standard ANOVA (analysis of variances) output including between-sample and intra-sample variances $S_{bsi}^2$ and $S_{isi}^2$ of the material inhomogeneity shown as $MS$ of

Rows and of Columns; $F$(Rows) is the Fisher's ratio characterizing between-sample inhomogeneity; $F$(Columns) is the Fisher's ratio characterizing intra-sample inhomogeneity; and $F_{crit}$ is the $F$ critical value at the level of confidence of 95%.

A range greater than the 95% limit is not observed. The $F$ values are less than the critical ones at the level of confidence of 95%. Therefore, the four samples are homogeneous in slump units. Since there are samples taken at the beginning, the middle and the end of the experiment, all the material produced for the experiment is also assessed as homogeneous in these units.

The assigned/certified slump value is calculated as the RL average result: $C_{cert} = \sum_{n=1}^{4} (X_{avg})_n / 4 = 111.9$ mm, where n is the number of the samples tested by RL. The standard uncertainty of the assigned value is $u = [u_{mRL}^2 + S_{bsi}^2 + S_{isi}^2/2]^{1/2} = 5.5$ mm.

#### *Analysis of the results obtained by the PT participants*

The participant results of slump determination are presented in Table 3, where $u_{mLP}$ is the standard measurement/test uncertainty declared by the participant; $L = 2.77 u_{mLP}$ is the limit of the range of the participant duplicates at the level of confidence of 0.95; $B = X_{avg} - C_{cert}$ is the bias of the average participant result from the assigned/certified value; $u_{comb} = (u_{mLP}^2 + u^2)^{1/2}$ is the combined standard test uncertainty; and $\zeta = B/u_{comb}$ is the *zeta* score.

The ranges of duplicates of laboratories No. 1, 2, 15 and 16 (shown by bold in Table 3) are greater than their 95% limit. It means that either the slump determinations in these cases were performed not completely by the corresponding laboratory SOP, or the measurement/test uncer-

**Table 2** RL results of the IHRM study in slump units

| Sample number | Slump, mm | | $R$, mm | $L$, mm | $X_{avg}$, mm | |
| | $X_1$ | $X_2$ | | | | |
|---|---|---|---|---|---|---|
| 1 | 105 | 110 | 5 | 8 | 107.5 | |
| 12 | 110 | 115 | 5 | 8 | 112.5 | |
| 23 | 115 | 115 | 0 | 8 | 115.0 | |
| 29 | 115 | 110 | 5 | 8 | 112.5 | |
| ANOVA | | | | | | |
| *Summary* | *Count* | *Sum* | *Average* | *Variance* | | |
| Row 1 | 2 | 215 | 107.5 | 12.5 | | |
| Row 2 | 2 | 225 | 112.5 | 12.5 | | |
| Row 3 | 2 | 230 | 115 | 0 | | |
| Row 4 | 2 | 225 | 112.5 | 12.5 | | |
| Column 1 | 4 | 445 | 111.25 | 22.917 | | |
| Column 2 | 4 | 450 | 112.5 | 8.333 | | |
| *Source of variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F* crit |
| Rows | 59.375 | 3 | 19.792 | 1.727 | 0.332 | 9.277 |
| Columns | 3.125 | 1 | 3.125 | 0.273 | 0.638 | 10.128 |
| Error | 34.375 | 3 | 11.458 | | | |
| Total | 96.875 | 7 | | | | |

**Table 3** Results of slump determination obtained by the PT participants

| Lab j | Slump, mm | | $u_{mLP}$, mm | $R$, mm | $L$, mm | $X_{avg}$, mm | $B$, mm | $u_{comb}$, mm | $\zeta$ |
|---|---|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | | | | | | | |
| 1 | 115 | 135 | 6.3 | **20** | 17 | 125.0 | 13.1 | 8.4 | 1.6 |
| 2 | 135 | 115 | 6 | **20** | 17 | 125.0 | 13.1 | 8.1 | 1.6 |
| 3 | 50 | 55 | 6.3 | 5 | 17 | 52.5 | − 59.4 | 8.4 | *− 7.1* |
| 4 | 135 | 140 | 4 | 5 | 11 | 137.5 | 25.6 | 6.8 | *3.8* |
| 5 | 145 | 140 | 3.3 | 5 | 9 | 142.5 | 30.6 | 6.4 | *4.8* |
| 6 | 110 | 100 | 6 | 10 | 17 | 105.0 | − 6.9 | 8.1 | − 0.8 |
| 7 | 100 | 115 | 16.1 | 15 | 45 | 107.5 | − 4.4 | 17.0 | − 0.3 |
| 8 | 125 | 125 | 9.1 | 0 | 25 | 125.0 | 13.1 | 10.6 | 1.2 |
| 9 | 110 | 125 | 6.3 | 15 | 17 | 117.5 | 5.6 | 8.4 | 0.7 |
| 10 | 105 | 110 | 3.5 | 5 | 10 | 107.5 | − 4.4 | 6.5 | − 0.7 |
| 11 | 140 | 145 | 3.3 | 5 | 9 | 142.5 | 30.6 | 6.4 | *4.8* |
| 12 | 125 | 110 | 6 | 15 | 17 | 117.5 | 5.6 | 8.1 | 0.7 |
| 13 | 150 | 140 | 4 | 10 | 11 | 145.0 | 33.1 | 6.8 | *4.9* |
| 14 | 135 | 140 | 4 | 5 | 11 | 137.5 | 25.6 | 6.8 | *3.8* |
| 15 | 150 | 125 | 6 | **25** | 17 | 137.5 | 25.6 | 8.1 | *3.1* |
| 16 | 160 | 140 | 6.3 | **20** | 17 | 150.0 | 38.1 | 8.4 | *4.6* |
| 17 | 140 | 135 | 4 | 5 | 11 | 137.5 | 25.6 | 6.8 | *3.8* |
| 18 | 90 | 100 | 6.3 | 10 | 17 | 95.0 | − 16.9 | 8.4 | − 2.0 |
| 19 | 120 | 100 | 16.5 | 20 | 46 | 110 | − 1.9 | 17.4 | − 0.1 |
| 20 | 110 | 115 | 6 | 5 | 17 | 112.5 | 0.6 | 8.1 | 0.1 |
| 21 | 90 | 95 | 6 | 5 | 17 | 92.5 | − 19.4 | 8.1 | *− 2.4* |
| 22 | 100 | 125 | 16.9 | 25 | 47 | 112.5 | 0.6 | 17.8 | 0.0 |
| 23 | 115 | 110 | 6 | 5 | 17 | 112.5 | 0.6 | 8.1 | 0.1 |
| 24 | 115 | 125 | 6 | 10 | 17 | 120.0 | 8.1 | 8.1 | 1.0 |
| 25 | 100 | 95 | 6 | 5 | 17 | 97.5 | − 14.4 | 8.1 | − 1.8 |

tainty $u_{mLP}$ declared in the SOP is evaluated not adequately enough (lower than in reality). The unsatisfactory score values $|\zeta| > 3$ of laboratories No. 3–5, 11, 13–17, as well as the questionable score value $|\zeta| > 2$ of laboratory No. 21 are italics in Table 3. The problem is that measurements of heights of the slump- and concrete cones, i.e. the length measurements (with a ruler having subdivisions of 1 mm) are not the limiting/dominant stage of the test. Manual filling the slump-cone with concrete using a steel rod, as well as the cone careful manual lifting during 5–10 s are a kind of art. Naturally, technicians perform these operations not equally (depending on their experience, temperament, etc.), especially in different laboratories. However, Table 3 contains the score values individual for every participant, while assessment of the results comparability (as a group) can be helpful.

*Comparability assessment*

The hypothesis about normal distribution of 58 slump determination results $X$ ($4 \times 2 = 8$ RL results and $25 \times 2 = 50$ participant results) is not rejected according to the Cramer von Mises criterion: the empirical value $\omega^2 = 0.88$ is less than the critical value of 2.50 at the level of confidence of 95%. The $X$ histogram and the fitted normal
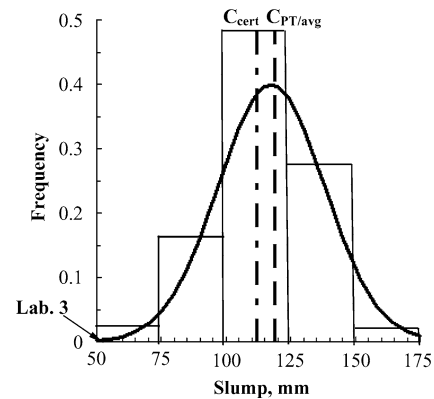


**Fig. 1** Histogram and fitted distribution of slump determination results. $C_{cert}$ is the IHRM assigned/certified slump value, and $C_{PT/avg}$ is the average slump value obtained by the PT participants. The pointer shows the average slump result obtained by laboratory (participant) No. 3

distribution are presented in Fig. 1. In this case the distribution of $X_{avg}$ values is also normal by the central limit theorem [4]. The total average result of the PT participants $C_{PT/avg} = \sum_{j=1}^{N} (X_{avg})_j / 25 = 118.6$ mm is shown in Fig. 1 by a dotted line. The standard deviation of $X_{avg}$ values from $C_{PT/avg}$ is $S_{PT} = 21.5$ mm. The assigned/certified

**Table 4** RL results of the IHRM study in units of compressive strength

| Sample number | Strength, MPa | | | | | | $R$, MPa | $L$, MPa | $X_{\mathrm{avg}}$, MPa |
|---|---|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | | | |
| 1 | 32 | 34 | 33.5 | 32.5 | 32.5 | 32.5 | 1.5 | 4 | 32.83 |
| 12 | 31 | 31.5 | 30.5 | 31.5 | 31.5 | 32 | 1.5 | 4 | 31.33 |
| 23 | 32 | 32.5 | 32 | 31.5 | 31.5 | 32 | 1 | 4 | 31.92 |
| 29 | 31.5 | 32.5 | 32 | 32 | 31.5 | 31.5 | 1 | 4 | 31.83 |
| ANOVA | | | | | | | | | |
| *Summary* | *Count* | *Sum* | *Average* | *Variance* | | | | | |
| Row 1 | 6 | 197 | 32.833 | 0.567 | | | | | |
| Row 2 | 6 | 188 | 31.333 | 0.267 | | | | | |
| Row 3 | 6 | 191.5 | 31.917 | 0.142 | | | | | |
| Row 4 | 6 | 191 | 31.833 | 0.167 | | | | | |
| Column 1 | 4 | 126.5 | 31.625 | 0.229 | | | | | |
| Column 2 | 4 | 130.5 | 32.625 | 1.062 | | | | | |
| Column 3 | 4 | 128 | 32 | 1.5 | | | | | |
| Column 4 | 4 | 127.5 | 31.875 | 0.229 | | | | | |
| Column 5 | 4 | 127 | 31.75 | 0.25 | | | | | |
| Column 6 | 4 | 128 | 32 | 0.167 | | | | | |
| Source of variation | SS | *df* | MS | $F$ | *P*-value | $F_{\mathrm{crit}}$ | | | |
| Rows | 7.031 | 3 | 2.344 | 10.714 | 0.001 | 3.287 | | | |
| Columns | 2.427 | 5 | 0.485 | 2.219 | 0.106 | 2.901 | | | |
| Error | 3.281 | 15 | 0.219 | | | | | | |
| *Total* | 12.740 | 23 | | | | | | | |

value $C_{\mathrm{cert}} = 111.9$ mm is also shown in Fig. 1 by another dotted line.

Values $u$ and $S_{\mathrm{PT}}$ are statistical sample estimates of population values $\sigma_{\mathrm{cert}}$ and $\sigma_{\mathrm{PT}}$ allowing to assume the ratio $\gamma = \sigma_{\mathrm{cert}}/\sigma_{\mathrm{PT}} = 0.5$ for the slump determinations. Since $|C_{\mathrm{cert}} - C_{\mathrm{PT/avg}}|/S_{\mathrm{PT}} = 0.31 < 0.40$, one can say that comparability of the results of the 25 PT participants is satisfactory at the level of confidence of 95% [3]. It means that customers can use results of the PT participants as comparable ones. Even the outlying average result of laboratory No. 3 shown in Fig. 1 by a pointer is within the distribution framework. This conclusion contradicts to a number of unsuccessful score values in Table 3. It probably indicates necessity of further improving uncertainty evaluation for correct interpretation of slump determination results.

Cone height/length measurements are traceable to the known measurement standards and the SI unit, while other important test stages (the slump-cone filling and lifting) are not traceable. It is similar to analyte extraction in chemical analytical methods, where recovery can dramatically influence the analysis results, even if the measurement/detection of amount of substance at the final stages of the analysis is precise and traceable. Internationally accepted RMs allow to overcome this problem in chemistry. However, the concrete IHRM developed for our experiment is by definition unstable and cannot be used elsewhere. Therefore, the comparability discussed here is of local relevance only.

Compressive strength determination

*Homogeneity estimation, assigned value and its uncertainty*

The RL results of strength determination on the 28th day obtained using a vibrating table for the test cube preparation, are presented in Table 4. The rows in the table represent samples; $X_1, X_2, \ldots, X_6$ are the replicate values; $R = X_{\mathrm{max}} - X_{\mathrm{min}}$ is the range, $X_{\mathrm{max}}$ and $X_{\mathrm{min}}$ are the maximal and minimal of the six replicate values; $L = 4.03\, u_{\mathrm{mRL}}$ is the range limit at the level of confidence of 95%, $u_{\mathrm{mRL}} = 1$ MPa is the standard measurement/test uncertainty declared by RL; and $X_{\mathrm{avg}} = \sum_{i=1}^{6} X_i/6$ is the average result of the strength determinations for a sample. A standard ANOVA output is also attached, like in Table 2. The intra-sample inhomogeneity in strength units is statistically insignificant. However, the Fisher's ratio characterizing the between-sample inhomogeneity $F(\mathrm{Rows})$ is greater than the critical value $F_{\mathrm{crit}}$ at the level of confidence 95%, i.e. the IHRM between-sample inhomogeneity in strength units is statistically significant. To evaluate the inhomogeneity, the value $F(\mathrm{Rows})$ is recalculated as ratio of the between-sample variance $S_{\mathrm{bsi}}^2$ (shown as $MS$ (Rows) in the output) to the measurement/test uncertainty variance $u_{\mathrm{mRL}}^2$: $F = 2.344/1 = 2.344 < F_{\mathrm{crit}}$. Therefore, the inhomogeneity is not significant in comparison to the RL measurement/test uncertainty, and the IHRM can be used for the experiment.

**Table 5** Results of compressive strength determination obtained by the PT participants

| Lab $j$ | Cube prepar | Strength, MPa | | | | | | $u_{mLP}$, MPa | $R$, MPa | $L$, MPa | $X_{avg}$, MPa | $B$, MPa | $u_{comb}$, MPa | $\zeta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | | | | | | | |
| 1 | hand | 27.5 | 27.5 | 26 | 28.5 | 29 | 28 | 1.9 | 3 | 8 | 27.75 | −4.2 | 2.6 | −1.60 |
| 2 | vibr | 30.5 | 29 | 30.5 | 28.5 | 29 | 30 | 5.2 | 2 | 21 | 29.58 | −2.4 | 5.5 | −0.43 |
| 3 | hand | 33 | 33 | 33 | 33 | 33 | 33 | 1.9 | 0 | 8 | 33.00 | 1.0 | 2.6 | 0.39 |
| 4 | hand | 31.5 | 31.5 | 32 | 30 | 31 | 32 | 1 | 2 | 4 | 31.33 | −0.6 | 2.1 | −0.31 |
| 5 | hand | 30.5 | 31 | 30 | 30.5 | 31.5 | 30 | 0.67 | 1.5 | 3 | 30.58 | −1.4 | 2.0 | −0.71 |
| 6 | hand | 29.5 | 30 | 29 | 29 | 30 | 28.5 | 5.2 | 1.5 | 21 | 29.33 | −2.6 | 5.5 | −0.48 |
| 7 | hand | 29 | 27 | 29.5 | 31.5 | 27.5 | 29.5 | 2.5 | 4.5 | 10 | 29.00 | −3.0 | 3.1 | −0.97 |
| 8 | hand | 27.5 | 27 | 26.5 | 27.5 | 27.5 | 27.5 | 2 | 1 | 8 | 27.25 | −4.7 | 2.7 | −1.74 |
| 9 | hand | 30.5 | 30 | 30.5 | 29 | 30.5 | 29.5 | 1.9 | 1.5 | 8 | 30.00 | −2.0 | 2.6 | −0.75 |
| 10 | vibr | 34 | 33 | 34 | 33 | 33.5 | 33 | 0.41 | 1 | 2 | 33.42 | 1.4 | 1.9 | 0.76 |
| 11 | hand | 30 | 29.5 | 28.5 | 30.5 | 29 | 30.5 | 0.67 | 2 | 3 | 29.67 | −2.3 | 2.0 | −1.17 |
| 12 | hand | 31 | 31 | 31 | 30.5 | 31 | 30 | 5.2 | 1 | 21 | 30.75 | −1.2 | 5.5 | −0.22 |
| 13 | hand | 28 | 27.5 | 27 | 28 | 28.5 | 29 | 1 | 2 | 4 | 28.00 | −4.0 | 2.1 | −1.89 |
| 14 | hand | 32 | 30.5 | 31.5 | 32 | 30.5 | 31 | 1 | 1.5 | 4 | 31.25 | −0.7 | 2.1 | −0.35 |
| 15 | hand | 33.5 | 33 | 33 | 33 | 32.5 | 32 | 5.2 | 1.5 | 21 | 32.83 | 0.9 | 5.5 | 0.15 |
| 16 | hand | 26.5 | 20.5 | 27 | 27 | 27.5 | 27.5 | 1.9 | 7 | 8 | 26.00 | −6.0 | 2.6 | −2.26 |
| 17 | hand | 31 | 30 | 29.5 | 28.5 | 29.5 | 29.5 | 1 | 2.5 | 4 | 29.67 | −2.3 | 2.1 | −1.10 |
| 18 | hand | 31 | 30 | 30 | 31 | 30.5 | 30 | 1.9 | 1 | 8 | 30.42 | −1.6 | 2.6 | −0.59 |
| 19 | hand | 30 | 30 | 29.5 | 28.5 | 29.5 | 28.5 | 2.5 | 1.5 | 10 | 29.33 | −2.6 | 3.1 | −0.85 |
| 20 | hand | 32.5 | 32.5 | 31 | 32 | 31 | 31.5 | 5.2 | 1.5 | 21 | 31.75 | −0.2 | 5.5 | −0.04 |
| 21 | hand | 30 | 30 | 29.5 | 30.5 | 30 | 30.5 | 5.2 | 1 | 21 | 30.08 | −1.9 | 5.5 | −0.34 |
| 22 | hand | 27 | 28 | 28 | 28.5 | 27.5 | 28 | 2.4 | 1.5 | 10 | 27.83 | −4.1 | 3.0 | −1.38 |
| 23 | hand | 32 | 31.5 | 31.5 | 32 | 32 | 31 | 5.2 | 1 | 21 | 31.67 | −0.3 | 5.5 | −0.06 |
| 24 | hand | 32 | 31 | 31 | 31.5 | 31 | 31 | 5.2 | 1 | 21 | 31.25 | −0.7 | 5.5 | −0.13 |
| 25 | hand | 30.5 | 33 | 31.5 | 33 | 31.5 | 33.5 | 5.2 | 3 | 21 | 32.17 | 0.2 | 5.5 | 0.03 |

Thus, both intra- and between-sample inhomogeneity components are taken into account, like in the slump study. Hence, the assigned/certified strength value is $C_{cert} = \sum_{n=1}^{4} (X_{avg})_n/4 = 32.0$ MPa, and its standard uncertainty is $u = [u_{mRL}^2 + S_{bsi}^2 + S_{isi}^2/6]^{1/2} = 1.9$ MPa.

*Analysis of the results obtained by the PT participants*

The participant results of strength determination obtained on the 28th day are presented in Table 5. All the results, both with hand test cube preparation ("hand" in Table 5) and using a vibrating table ("vibr") are satisfactory concerning their ranges and ζ-score values. The only questionable score value is $2 < |\zeta = -2.26| < 3$ obtained by laboratory No. 16 (*italics* in Table 5). Probably it is a random deviation: five laboratories out of 100, i.e. one out of 20 laboratories, can have $|\zeta| > 2$ at the level of confidence of 95%.

*Comparability assessment*

The hypothesis about normal distribution of 174 strength determination results $X$ ($4 \times 6 = 24$ RL results and $25 \times 6 = 150$ participant results) is not rejected according to the Cramer von Mises criterion: the empirical value $\omega^2 =$
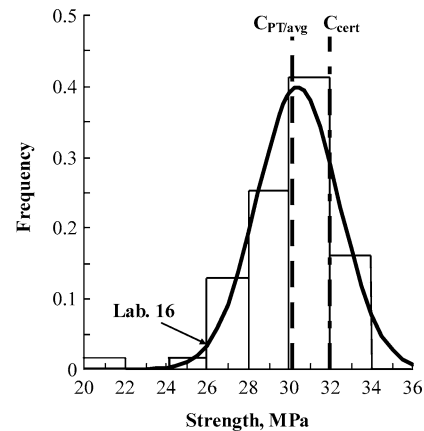
**Fig. 2** Histogram and fitted distribution of compressive strength determination results. $C_{cert}$ is the IHRM assigned/certified strength value, and $C_{PT/avg}$ is the average strength value obtained by the PT participants. The pointer shows the average slump result obtained by laboratory (participant) No. 16

1.53 is less than the critical value of 2.50 at the level of confidence of 95%. The $X$ histogram and the fitted normal distribution are presented in Fig. 2. The total average result of the participants $C_{PT/avg} = \sum_{j=1}^{N} (X_{avg})_j/25 = 30.2$ MPa is shown in Fig. 2 by a dotted line. The standard deviation of $X_{avg}$ from $C_{PT/avg}$ is $S_{PT} = 1.9$ MPa. The assigned/certified

value $C_{cert} = 32.0$ MPa is shown in Fig. 2 by another dotted line, like in Fig. 1. The average result obtained by laboratory No.16 having a questionable score value is shown in Fig. 2 by a pointer.

Calibrated testing machines used in RL and in the laboratories participated in PT allow to measure pressure with standard uncertainty of less than 2%. Therefore, the values $u$ and $S_{PT}$ are equal and the assumption $\gamma = 1$ is reasonable here. Since $|C_{cert} - C_{PT/avg}|/S_{PT} = 0.95 < 1.04$, comparability of the results of the 25 participants can be assessed as satisfactory at the level of confidence of 95% [3]. Thus, the individual score and the comparability (group) assessment coincide for strength determinations.

The pressure measurement applied to a test cube in order to destroy it is traceable to the corresponding international measurement standards. The stages of test cube preparation and hardening during 28 days can be performed in conditions controlled by using traceable measurements. They are less depending on the technician's art than in slump determination. Therefore, traceability of the IHRM assigned strength value to international measurement standards can be achieved theoretically. However, again, the concrete IHRM is intended for the single use only. Thus, the comparability discussed for strength determinations is also of local relevance.

## Conclusions

1. A PT scheme is developed for comparability assessment of results of concrete slump and compressive strength determination. The scheme is based on preparing of a concrete IHRM test portion/sample at a reference laboratory in the same conditions for every PT participant, and on using of the IHRM as a local working measurement standard.
2. The IHRM produced and studied for the PT in Israel is found homogeneous in both slump and strength units on the level of measurement/test uncertainties declared by the reference laboratory.
3. Since traceability of the concrete IHRM assigned values to international measurement standards and SI units cannot be stated, local comparability of the results is assessed. It is shown, that comparability of the slump and compressive strength determination results is satisfactory, while the uncertainty evaluation for slump results requires additional efforts.

## References

1. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML (2006) International Vocabulary of Basic and General Terms in Metrology. 3rd edn., ISO, Geneva
2. EURACHEM/CITAC Guide (2003) Traceability in Chemical Measurement. A Guide to Achieving Comparable Results in Chemical Measurement, EURACHEM/CITAC, www.eurachem.ul.pt
3. Kuselman I (2006) Accred Qual Assur 10:466–470
4. Kuselman I (2006) Accred Qual Assur 10:659–663
5. Kuselman I (2006) In the Proceedings of the International Workshop "Combining and reporting analytical results. The role of (metrological) traceability and (measurement) uncertainty for comparing analytical results". The Royal Society of Chemistry Special Series Proceedings Book, Cambridge, UK
6. CCRL. Concrete Proficiency Sample Program, www.ccrl.us
7. Youden WJ (1959) Proc Am Soc Test Mat 59:1120–1128
8. Israeli Standard 26, part 2 (1992) Methods of testing concrete: Qualities of fresh concrete
9. Israeli Standard 26, part 4 (1985) Methods of testing concrete: Strength of hardened concrete
10. Kuselman I, Pavlichenko M (2004) Accred Qual Accur 9:399–403
11. Kuselman I (2004) J Metrol Soc India 19(4):245–252
12. Israeli Standard 118 (2003) Concrete for structural uses: Production control and compressive strength