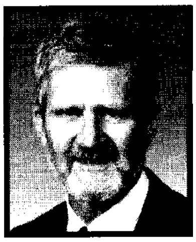
Inter-laboratory proficiency evaluations between several Gauteng-based concrete and aggregate testing laboratories

TECHNICAL PAPER
Journal of the South African Institution of Civil Engineering, 44(1) 2002,
Pages 31–34, Paper 489

G R H Grieve

The South African National Accreditation System requires accredited civil engineering laboratories to participate in inter-laboratory precision evaluations of the tests for which they are accredited. Results are presented for evaluation of a number of common aggregate and concrete tests conducted by seven laboratories which have participated in an evaluation programme for some time. The data were analysed and proposals have been made to assist civil engineering laboratories to assess the precision of their test results.



GRAHAM GRIEVE qualified as a civil engineer in 1970 at the University of Cape Town. His first job was with the Department Kustonderzoek, of Rijkswaterstaat (the equivalent of the Department of Water Affairs in SA) in the Netherlands for one year, before returning to South Africa to work at the Coastal Research Unit of the CSIR in Stellenbosch. He then broadened his experience by working as a resident engineer for B S Bergman & Partners on a road and bridge contract in Tzaneen in 1974. In 1978 he was transferred to the Pretoria office of B S Bergman and Partners with responsibility for materials testing and design of road projects.
He completed a BSc (Ing) (Hons) at Pretoria University in 1980 and an MIng degree in 1984. In 1991 he was awarded a PhD from the University of the Witwatersrand for research on practical use of fly-ash in concrete.
He joined the Portland Cement Institute (PCI) as Director of Laboratory Services in 1982 and was promoted to Deputy Executive Director of the Institute in 1987. He took over as Executive Director in May 1990.
As a representative of the cement, lime and concrete industries, he has been a member of the SABS ISO 9000 Advisory Board since its inception. At the beginning of 2000 he was elected chairman of this committee. He has been a member of various SABS cement aggregate and concrete technical committees since the early 1980s. Earlier this year he was made focus area champion of the Primary Construction Materials standards and codes (covering cement, lime, aggregates and the stabilisation of soils) under the responsibility of the SABS TC 5120.61 'Construction standards'.

INTRODUCTION

The South African National Accreditation System (SANAS) requires that laboratories accredited under SABS 0259:1990 'General Requirements for the Competence of Calibration and Testing Laboratories' should participate in regular inter-laboratory precision evaluations. The objective of such evaluations is to improve the reproducibility and repeatability of the tests being performed, and thereby ensure that the customers of such laboratories receive reliable results. This requirement will be reinforced when accredited laboratories convert to being accredited under ISO guide 17025; this process will be have to have been completed by the end of 2002.

A Concrete Sub-committee was formed under the Specialist Technical Committee (STC) of SANAS and the participating laboratories undertook a series of tests. The test data were submitted to a statistician recommended by SANAS, Mr W Breytenbach. This paper presents details of the tests undertaken and the interpretation of the data analysis. A suggestion is made at the end of the paper on the interpretation of test results reported by a laboratory.

PARTICIPATING ORGANISATIONS

The laboratories that participated in the testing programmes described in this report were as follows:

- Alpha Stone and Ready Mix
- Cement & Concrete Institute
- Civilab
- Concrete Testing Services
- Lafarge Ready Mix
- Lafarge Technical Services Department
- Slagment Pty Ltd

All these laboratories are either already accredited under SABS 0259 or are in the process of developing their quality systems to qualify for accreditation and all have an interest in achieving reliable results in each test performed.

TEST PROGRAMME

While the work undertaken by the group of seven laboratories has been under way for some time, the data that were analysed statistically were generated in 1998 and 1999.

The following tests were included in this evaluation:

- SABS Method 845-1994, Bulk densities and voids contents of aggregates, Pretoria: SABS, 1994.
- SABS Method 844-1994, Particle and relative density of aggregates, Pretoria: SABS, 1994.
- SABS Method 829-1994, Sieve analysis, fines content and dust content of aggregate, Pretoria: SABS, 1994.
- SABS Method 863-1994, Concrete tests - compressive strength of hardened concrete, Pretoria: SABS, 1994.
- SABS Method 865-1994, Concrete tests - the drilling, preparation, and testing for compressive strength of cores taken from hardened concrete, Pretoria: SABS, 1994.

ACCURACY, REPRODUCIBILITY AND REPEATABILITY

When the result of a measurement is considered, this is usually viewed as an individual event. The user unconsciously assumes that the result represents the true value of the characteristic measured. However, this single measurement originates from a chain of events, all full of imperfections and errors which shed doubt about the exactness of the result. The extent of this doubt should be determined, evaluated and controlled.

In this paper, the closeness to the true value is defined as accuracy. In many cases of civil engineering metrology, there may be no known true value for specific properties of the materials being tested. In such a situation, either a single test facility is accepted as reporting the true value – ie a reference laboratory – or the mean of the results from several competent laboratories can be taken as the true value.

The evaluation of accuracy should include factors such as the process of calibration and the control of accuracy between calibration intervals

Terms such as repeatability, reproducibility and random error refer to the closeness of agreement among repeated measurements made on the same specimen. This procedure considers accuracy as the higher order term, which includes repeatability and reproducibility.

Repeatability refers to the closeness of agreement when measurements are made under essentially the same conditions (same operator, conditions, locations, etc and within a short period of time). Reproducibility refers to the closeness of agreement when measuring under essentially changed or different conditions (different operator, conditions, locations, etc and over a longer time period).

As with accuracy, repeatability and reproducibility should be evaluated and controlled.

In order to evaluate repeatability and reproducibility, consensus should be reached on what the terms 'essentially the same' and 'essentially different' mean. A measurement system is a process consisting of different elements:

- Essentially the same conditions are created or simulated when repeated measurements are made on the same item or specimen over a short period of time without any deliberate changes. When evidence exists that the variation within these measurements is stable (or in statistical control) this 'within-group' variation can be used to represent the measurement system's repeatability.
- Essentially different conditions, on the other hand, are created or simulated when repeated measurements are made on the same specimen over a lengthy period of time that represents normal operating conditions. During normal operating conditions the different pieces of equipment and operators usually associated with that specific measurement system are then allowed to change as they normally would.
- Replacing one element with another, eg replacing the instrument, using another method or replacing the operator with a person not normally associated with being an operator in this measurement process is actually the establishment of a new measurement system.

Putting these concepts together, repeatability and reproducibility can thus be evaluated by taking small subgroups of repeated measurements under essentially the same conditions (giving within-group variation) over a period long enough to give between-subgroup variations. It is, however, essential that the same configuration is used, that is, the same instruments and other equipment, methods and operator(s) to represent all expected operational conditions normally associated with the measurement system. The 'within-group' variation will refer to repeatability, while the 'between-subgroup' can be used to estimate reproducibility. A valid statement of repeatability and reproducibility will therefore include a definition/description of the elements of the measurement system, which then will define what 'essentially the same conditions' implies.

The evaluation of repeatability and reproducibility should highlight the factors most likely to influence them. They include, but are not limited to, the following prominent influences:

- operator sensitivity, which refers to Man/Method/Equipment combinations giving different results when the same test specimen is measured repeatedly by different operators
- component sensitivity, which refers to variations in different test specimens being measured giving different results when measured by the same operator

INTERNATIONAL PRACTICE

Aggregate tests

A literature study was undertaken to determine the international approach to within-test and inter-laboratory precision. Many of the American Society for Testing and Materials (ASTM) test methods have precision estimates where the single operator and multi-laboratory standard deviation and range are provided. Tables 1 and 2 below provide values for some of the aggregate test methods.

Table 1 Published data on particle and bulk density in ASTM C29/C29M-97, Standard test method for bulk density ('unit weight') and voids in aggregates, for fine aggregates

Tests on fine aggregate	Repeatability		Reproducibility	
	Standard deviation	Range	Standard deviation	Range
Particle density (kg/m^3)	11	32	23	66
Bulk density (kg/m^3)	14	40	44	125

Table 2 Published data on particle and bulk density in ASTM C29/C29M-97, Standard test method for bulk density ('unit weight') and voids in aggregates, for coarse aggregates

Tests on coarse aggregate	Repeatability		Reproducibility	
	Standard deviation	Range	Standard deviation	Range
Particle density (kg/m^3)	9	25	13	38
Bulk density (kg/m^3)	14	40	30	85

In the BS 812 series of test on aggregates, the approach is to provide estimates of repeatability and reproducibility of the test method where this information is available. Table 3 below summarises some of these values.

Table 3 Published data on particle and bulk density in BS 812 Part 2-1995, Testing aggregates. Part 2. Methods for determination of density, for fine and coarse aggregates

Tests	Repeatability range	Reproducibility range
Particle density (kg/m^3)	20	40
Bulk density (kg/m^3)	10	20

For the purposes of recommending limits for intra- and inter-laboratory precision for South African aggregate testing laboratories the authors have relied on the ASTM philosophy for aggregate tests.

A test programme was arranged to determine the repeatability and reproducibility of the following tests:

- apparent relative density
- loose and consolidated bulk density
- sieve analysis

Each laboratory was required to carry out 20 repetitions of a particular test. Other laboratories undertook other tests. Where possible, the same test sample was used for all the tests. In the case of the sieve analysis on the fine aggregate sample, 20 sub-samples were riffled out from a single homogenised bulk sample; this permitted the 'destructive' process of washing to be repeated. The results were analysed independently and are reported below in tables 4 to 8.

Aggregate test results

Particle and relative density of aggregates to SABS Method 844-1994

Table 4 Relative density test results

Parameter	Sand	Stone
Mean value	2,987	2,941
Repeatability	0,012	0,031
Reproducibility	0,015	0,029

Typically the magnitude of repeatability would be expected to be smaller than that of reproducibility. In the tables above this is not always the case. It is believed that this can be ascribed to the probability that the measurement systems

Bulk densities on fine aggregates to SABS Method 845-1994

Table 5 Fine aggregate bulk density test results

Parameter	LBD	CBD
Mean value	1650	1870
Repeatability	26,9	9,3
Reproducibility	28,8	14,9

Bulk densities on coarse aggregates to SABS Method 845-1994

Table 6 Coarse aggregate bulk density test results

Parameter	LBD	CBD
Mean value	1330	1470
Repeatability	22,7	11,7
Reproducibility	18,5	17,0

Sieve analysis, fines content and dust content of fine aggregate to SABS Method 829-1994

Table 7 Results of sieve analyses on fine aggregate

	Percentage passing sieve size (mm)							
	6,7	4,75	2,36	1,18	0,6	0,3	0,15	0,075
Mean	100	98,8	91,8	68,9	34,8	4,4	0,7	0,3
Repeatability	–	0,87	1,02	2,46	2,20	0,90	0,68	0,64
Reproducibility	–	0,91	1,32	2,61	2,28	1,08	0,87	0,64

Sieve analysis, fines content and dust content of coarse aggregate to SABS Method 829-1994

Table 8 Results of sieve analyses on coarse aggregates

	26,50	19,00	13,20	9,50	6,70
Mean	100	86,0	17,4	1,2	0,6
Repeatability	–	5,13	6,19	0,79	0,53
Reproducibility	–	5,93	5,40	0,89	0,54

Table 9 Concrete tests – compressive strength of hardened concrete to SABS Method 863-1994 (cube tests) (The results reported below for each laboratory are the average of the six cubes received for testing in MPa)

		Laboratory						
		1	2	3	4	5	6	7
Low strength: 100-mm cubes	Mean (MPa)	17,2	18,5	18,2	18,1	17,2	17,9	17,4
	Standard deviation (MPa)	0,79	0,35	0,55	0,24	0,41	1,09	0,62
	Coefficient of variation (%)	4,6	1,9	3,0	1,3	2,4	6,1	3,6
Medium strength: 100-mm cubes	Mean (MPa)	36,4	38,4	37,2	38,0	36,9	36,1	37,1
	Standard deviation (MPa)	2,18	1,27	0,85	1,83	1,21	0,57	1,39
	Coefficient of variation (%)	6,0	3,3	2,3	4,8	3,3	1,6	3,7
High strength: 100-mm cubes	Mean (MPa)	77,2	78,7	75,4	74,7	75,4	75,1	72,6
	Standard deviation (MPa)	1,37	2,23	2,36	2,37	1,97	1,31	4,08
	Coefficient of variation (%)	1,8	2,8	3,1	3,2	2,6	1,7	5,6
High strength: 150-mm cubes	Mean (MPa)	59,8	60,9	58,6	56,3	57,8	56,6	57,3
	Standard deviation (MPa)	2,33	2,23	1,58	1,83	2,61	2,49	1,60
	Coefficient of variation (%)	3,9	3,7	2,7	3,3	4,5	4,4	2,8

Table 10 Concrete tests – the testing for compressive strength of cores taken from hardened concrete to SABS Method 865-1994 (MPa)

	Set 1	Set 2	Set 3	Set 4	Set 5
Lab 1	28,8	31,7	34,5	32,2	30,7
Lab 2	30,5	27,5	34,7	30,7	32,2
Lab 3	33,0	28,8	31,0	29,2	29,0
Lab 4	32,5	28,7	33,0	30,0	29,7
Lab 5	32,5	28,7	33,8	–	–
Mean	31,47	9,07	33,40	30,5	30,29
Reproducibility	3,31	6,14	4,02	2,66	–

are in control to the extent that the variation observed when measurements are made under essentially different conditions are as small as the variations observed when measurements are made under essentially the same conditions.

Concrete tests

For the purposes of recommending limits for intra- and inter-laboratory precision for South African concrete testing laboratories the authors have relied on the BS 1881 Part 127 philosophy for the concrete compressive strength tests.

Two types of concrete test were carried out by some of the participating laboratories as follows:

- Verification of the compressive testing machine was carried out using the cube testing procedure described in the BS 1881 Part 127. In this case, for each strength range required, a team in the C&CI laboratory produced a single large batch of concrete. Representative cubes were compacted and cured for 28 days in the C&CI laboratory. Sets of six cubes were then distributed to each participating laboratory for testing on a particular date.
- Concrete cores were tested to evaluate the reproducibility and repeatability of the laboratory procedures for core preparation and testing. In the case of the core tests, cube specimens were compacted and cured for 28 days in one laboratory. Core specimens were then drilled horizontally from these cores and sets of three cores were distributed to each participating laboratory for testing on a particular date. Each participating laboratory was required to trim, cap, cure (ie soak for the required 48 hours) and carry out the compressive strength test on the cores.

Concrete test results

The results of these tests are provided in tables 9 and 10.

The sets of core specimens referred to in table 10 were made as described above. The result reported in table 10 was the mean of the compressive strength results of the three cores tested by each laboratory. Sets 1 to 5 were made at different times over the period between the beginning of 1999 and the end of 2000. All the 150-mm cubes from which the approximately 90-mm diameter cores were drilled were cast and cured at the C&CI laboratories. The coring was undertaken by Concrete Testing Services.

DISCUSSION

Although this trend is not shown in the results presented here, it should be noted that after the inter-laboratory testing was initiated, the gap between results from the different laboratories started closing very quickly. The range between highest and lowest results reported by different laboratories became smaller with repeated inter-laboratory proficiency testing. Such improvements in inter-laboratory precision result from the investigations and correc-

Table 11 Grading analysis recommendations

Coarse		Fine	
% Passing	Range	% Passing	Range
100–95	1,0	100–95	0,6
95–85	3,9	95–60	2,2
85–80	5,4	60–20	4,0
80–60	8,0	20–15	3,1
60–20	5,6	15–10	2,1
20–15	4,5	10–2	1,8
15–10	4,2	2–0	0,9
10–5	3,4		
5–2	3,0		
2–0	1,3		

tive actions by participating laboratories where their results are seen to differ significantly from those reported by other participants. It would probably be reasonable to expect that the same would happen for the results from any new laboratory that chose to participate in future inter-laboratory testing programmes.

The test data presented in the tables above generally indicate that the participating laboratories will report similar values for the type of material or compressive strength range of concrete measured in the tests included in the programme. It does not necessarily indicate that the inter-laboratory precision of the same tests on different materials or grades of concrete would be equally 'in control'. It is suggested, however, that the probability that this is the case is greater than for a laboratory which has not previously participated in such precision evaluations. These precision evaluations are part of a quality management process to which each of the participating laboratories has committed itself.

The statistical analysis of these data equips the participating laboratories to

- set internal control parameters for acceptance of test results, ie limits can be set within which repeat results should lie to be regarded as an acceptable set of results. If repeat results were to fall outside the set control limits, then this would represent an 'out of control' event which would require investigation, and where possible, corrective action. The test would also have to be repeated until the repeat result showed the test to be 'in control' again
- provide an assurance to the client that, should another of the participating laboratories be asked to undertake a duplicate test, the new result should be within a known small range of the first test result. This information would assist the client to make informed decisions about the acceptance of the product from which the test samples were taken

RECOMMENDED CONTROL LIMITS FOR REPEATABILITY

Evaluation of cube results (to BS 1881 Part 127)

The calculations to evaluate the cube results will be undertaken as follows:

(i) Calculation of arithmetic mean (ξ) and standard deviation (S) for each set.

(ii) Calculation of the difference d between a result and that of the reference machine from the equation

$$d = 100*(\xi - \xi_r)/ \xi_r \qquad (1)$$

where ξ_r is the mean from the machine considered to be the reference machine

(iii) Calculation of e from the equation

$$e = (91* \sqrt{S^2 + S_r^2})/ \xi_r \qquad (2)$$

(iv) Calculation of the 95% confidence limits

Lower confidence limit L = d – e (3)

Upper confidence limit U = d + e (4)

Interpretation of results:

(i) Assessment of difference in means

A result fails if L >0,0 or d >4,0

Or if U <0,0 or d <-4,0

(ii) Assessment of standard deviation

High strength. A result fails if S >3,0 and S >3,3 Sr

Low strength. A result fails if S >0,7 and S >3,3 Sr

The above evaluation procedure confirms whether or not the mean value of results generated by a specific testing machine lie within ± 4% of the mean from the 'reference machine' at a 95% confidence level. Where the value of 'd' for a specific machine lies outside ± 4% of the mean from the 'reference machine' this indicates an out-of-compliance situation.

Values 'L' and 'U' are calculations for the confidence limits based on the Student's 't' distribution.

In the event of a non-compliance, the suspect machine should be serviced or checked after which a further comparative test should be undertaken to confirm compliance.

EVALUATION OF AGGREGATE RESULTS

Based on ASTM recommendations and our own findings, the following tolerances are currently applied by the participating laboratories:

Bulk densities

+ and – 40 kg of mean value (coarse)
+ and – 60 kg of mean value (fine)

Relative density

+ and – 0,03 of mean value (coarse and fine)

CONCLUSIONS

Utilising the data presented in this paper, laboratories should be in a position to evaluate the variability of results obtained in inter- or intra-laboratory testing against the recommended tolerances shown above. This will assist in detecting 'out-of-control' events which should be investigated and subjected to the necessary corrective action. Continuous application of this principle should help laboratories to improve inter- and intra-laboratory testing precision.

It should be noted that the data submitted in this paper represents a snapshot of the situation at the time. Should any laboratory report results that are significantly different to those reported by the others (say differing by more than 20% from the mean of all the results) then this would represent an 'out-of-control' situation. Corrective action would be required and the effectiveness of the completed corrective action would need to be verified with at least one other of the participating laboratories. During the period covered by the results presented above all the laboratories appeared to be 'in control'.

A prospective consumer of testing services should confirm that the service provider selected is able to demonstrate that the test facility remains 'in control' for the required tests.

Acknowledgements

The assistance of the following persons is gratefully acknowledged in the preparation of this paper: Mr W Breytenbach, the statistician who analysed the data, and Messrs J Kellerman, D Tite, H Greyling and S Hafjee, who commented on and added to the draft paper.

References

ASTM C29/C29M-97. *Standard test method for bulk density ('unit weight') and voids in aggregates.* West Conshohocken: ASTM.

Breytenbach, W P 1999. Private communication on statistical analysis of test results. Johannesburg, December.

Breytenbach, W P 2000. *C&CI report – the analysis of Fosroc and random sample failure stress results.* Pretoria: Q&U cc.

BS 812 Part 2-1995. *Testing aggregates, Part 2, Methods for determination of density.* London: BSI.

BS 1881 Part 127-1990, *Testing concrete, Part 127, Method of verifying the performance of a concrete cube compression machine using the comparative cube test.* London: BSI.

SABS 0259-1990. *Code of practice for general requirements for the competence of calibration and testing laboratories.* Pretoria: SABS.

SABS Method 844-1994. *Particle and relative density of aggregates.* Pretoria: SABS.

SABS Method 845-1994. *Bulk densities and voids contents of aggregates.* Pretoria: SABS.

SABS Method 829-1994. *Sieve analysis, fines content and dust content of aggregate.* Pretoria: SABS.

SABS Method 863-1994. *Concrete tests – compressive strength of hardened concrete.* Pretoria: SABS.

SABS Method 865-1994. *Concrete tests – the drilling, preparation, and testing for compressive strength of cores taken from hardened concrete.* Pretoria: SABS.