Authorized Reprint from the Copyrighted Proceedings of the
AMERICAN SOCIETY FOR TESTING MATERIALS
Philadelphia, Pa.
Volume 59, 1959

STATISTICAL EVALUATION OF INTERLABORATORY CEMENT TESTS*

BY J. R. CRANDALL[1] AND R. L. BLAINE[1]

SYNOPSIS

A cooperative series of physical and chemical tests was made by 103 laboratories on 12 samples of cement during the period of 1 yr. Results of tests by all laboratories on each property were plotted in scatter-diagrams according to the Youden method to enable the participating laboratories to evaluate their results quickly. The results were treated statistically. Both statistical and graphic methods were employed to indicate the precision of the test methods and the extent of laboratory bias. The lack of precision with some of the tests indicated a need for improved test methods.

The work of each of the various laboratories was evaluated by a rating system. A large number of laboratories obtained results that were in good agreement, but a few laboratories showed poor agreement with the larger group. Many of the discrepant laboratories did not improve during the 1-yr period the test program was active.

Discrepant results obtained by a few laboratories inflated appreciably the standard deviations among laboratories for the various tests.

The need for a continuing reference sample program was indicated.

The values obtained in the testing of cement, as those obtained in testing of most materials, are subject to many variables. The methods of performing the tests are carefully specified as well as the requirements and tolerances for test equipment to be used. Cement testing laboratories have, in addition, had a unique service for the past 30 years in that the Cement Reference Laboratory (CRL) has inspected the equipment and instructed the personnel of these laboratories at regular intervals. A number of comparative samples have also been distributed by this group during the past 30 years. A report of the work done by the CRL and of the resulting improvements in test equipment was presented by J. R. Dise.[2]

The service of the CRL of inspecting laboratory equipment, instructing personnel and demonstrating test methods is costly; furthermore, because turnovers occur in testing personnel and changes occur in test methods and equipment, more frequent inspections would be desirable.

The CRL comparative tests as well as other interlaboratory tests have indicated that there are still many and large discrepancies between the results as reported by different laboratories. It has not been clear from previous comparative tests if these discrepancies resulted from normal variations to be expected or if

* Presented at a special session on Statistical Treatment of Interlaboratory Test Results, Including a New Graphical Method sponsored by Committee C-1 on Cement held on June 25, 1959 during the Sixty-second Annual Meeting of the Society.

[1] Concreting Materials Section, National Bureau of Standards, Washington, D. C.

[2] J. R. Dise, "Cement Reference Laboratory (1929–1959)," see p. 369, this publication.

laboratory bias was involved to any appreciable extent.

In order to determine the reproducibility of the various tests for cement now in use and to evaluate the nature of the discrepancies, as well as to offer a means for supplementing the work of the CRL, a Cement Reference Sample program was proposed at the 1957 ASTM Annual Meeting, to the Subcommittee on the CRL of Committee C-1, by the personnel of the National Bureau of Standards (NBS).

The Reference Sample Program was considered an extension of the NBS Standard Sample Program which has served laboratories everywhere for almost 60 yr. Before these standard samples of cement could be issued, however, it was necessary to solve problems relative to blending, packaging and distribution of the samples as well as tabulation, evaluation and presentation of the test results. Although the planned program of interlaboratory tests was quite extensive, it was necessary to limit the number of participating laboratories because of the present limited facilities for blending very large quantities of materials.

TABLE I.—MANUFACTURERS OF CEMENTS USED.

Universal Atlas Cement Co.
North American Cement Corp.
Standard Lime & Cement Co.
Lehigh Portland Cement Co.
Alpha Portland Cement Co.
Green Bay Division, Pittsburgh Coke & Chemical Co.
Lone Star Cement Co.
Medusa Portland Cement Co.

PARTICIPATING LABORATORIES

It was desirable in a program of this kind to have as broad a distribution of laboratories as possible, both with respect to interest and geography. Included in the program were both research and mill-control laboratories of cement manufacturers, State Highway laboratories, Federal laboratories, and commercial laboratories. Included among the cement manufacturer laboratories were larger companies who regularly conduct interlaboratory tests and smaller companies who have only one or two mills and whose opportunities for interlaboratory tests ordinarily are limited. All participating laboratories had been inspected by the CRL. There were 57 cement manufacturer laboratories, 24 State Highway laboratories, 11 commercial laboratories and 11 Federal Government laboratories, making a total of 103 laboratories from all sections of the country. A list of the participating laboratories is given in Appendix I.

TABLE II.—IDENTIFICATION OF PAIRS OF CEMENT SAMPLES.

Sample	Type	Lot Number	Date Shipped to Laboratories
No. 1	I	1	March 1958
No. 2	I	2	
No. 3	I	3	May 1958
No. 4	IA	4	
No. 5	II	5	July 1958
No. 6	IS	6	
No. 7	III	7	Sept. 1958
No. 8	IA	8	
No. 9	II	9	Nov. 1958
No. 10	V	10	
No. 11	IA	4	Jan. 1959
No. 12	I	3	

TEST PROGRAM

The test program consisted of physical and chemical tests by the 103 laboratories on 12 cement samples. Ten lots of different cements were used to prepare the 12 samples.

Nine of the ten cements used in the program were purchased from suppliers in the Washington, D. C., area, and one cement was donated by the manufacturer (Table I). Each lot of cement was from a single carload shipment. Six types of cement were used (Table II). Each lot consisted of 16 or 18 bags of cement, except lots 3 and 4, which were double that amount. Each cement was sieved through a No. 20 vibrating screen and each lot then blended for 2 to 3 hr in a 20-cu ft Patterson-Kelly blender equipped with an intensifier.

The cement was then packaged in plastic-lined, canvas bags, approximately 12 lb per package, and stored in sealed steel drums until shipment, except in the case of samples Nos. 11 and 12, where one half of each set of samples was stored on a skid in a storeroom. Tests for uniformity were made on every tenth sample bag packaged. Four separate test samples were taken from each of these bags, two being used for air permeability fineness tests and the other two for SO_3 determinations. Each sample was coded so that the test operator did not know which of the samples were duplicates from the same bag.

Uniformity tests were made comparing the differences between duplicate determinations on the same samples, the sums of duplicates, and the successive differences between sums. All these tests for uniformity were made in the NBS Washington laboratory, except tests on one pair of samples which were made in the NBS Seattle laboratory. In each instance, the statistical analysis indicated that the cement was well-blended.

ASTM test methods were used throughout the program, except as indicated in Table III. For samples Nos. 9 to 12, ASTM referee test methods were used for the analytical determinations, with Federal Specifications for the alkalies. In each case, the laboratories were requested to report results of single determinations.

Samples of two of the cements were supplied to the participating laboratories at 2-month intervals for a period of 1 yr.

TABLE III.—TEST METHODS USED.

Normal consistency of hydraulic cement	ASTM Method C 187 - 58[b]
Soundness, autoclave	ASTM Method C 151 - 58[b]
Time of setting, Gillmore	ASTM Method C 266 - 58 T[b]
Air content	ASTM Method C 185 - 58 T[b]
Compressive strength (6-cube batch)	ASTM Method C 109 - 58[b]
Fineness, air permeability	ASTM Method C 204 - 55[b]
Tensile strength[a]	ASTM Method C 190 - 58[b]
Fineness turbidimeter[a]	ASTM Method C 115 - 58[b]
Heat of hydration of portland cement[a]	ASTM Method C 186 - 55[b]
Time of setting, Vicat needle[a]	ASTM Method C 191 - 58[b]
False set[a]	ASTM Method C 359 - 56 T[b]
Sulfate resistance expansion[a]	Experimental Test

Test Procedure	ASTM Methods	Section of the Method Used for Samples Nos. 1 to 8	Section of the Method Used for Samples Nos. 9 to 12
Silicon dioxide (SiO_2)	C 114 - 58[b]	33	8
Aluminum oxide (Al_2O_3)	C 114 - 58[b]	12	12
Ferric oxide (Fe_2O_3)	C 114 - 58[b]	10 and 11	10 and 11
Calcium oxide (CaO)	C 114 - 58[b]	34 and 35	13
Magnesium oxide (MgO)	C 114 - 58[b]	36	14 and 15
Sulfur trioxide (SO_3)	C 114 - 58[b]	16	16
Loss on ignition	C 114 - 58[b]	20	20
Insoluble residue	C 114 - 58[b]	28	28
Sodium oxide (Na_2O)	C 114 - 58 T[b]	15 through 18	[c]
Potassium oxide (K_2O)	C 114 - 58 T[b]	15 through 18	[c]
Manganic oxide (Mn_2O_3)[a]	C 114 - 58[b]	49 and 50	...
Sulfide sulfur[a]	C 114 - 58[b]	17 through 19	...

[a] These properties determined on one pair of samples only.
[b] 1958 Book of ASTM Standards, Part 4.
[c] A Proposed Interim Federal Specification for Cement.

TABLE IV.—STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR DATA REPORTED BY ALL LABORATORIES.

Sample	No. 1		No. 2		No. 3		No. 4		No. 5		No. 6	
	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.
Water consistency, per cent	0.45	1.8	0.86	3.4	0.50	1.9	0.58	2.2	0.71	2.9	1.2	4.5
Initial setting, Gillmore, hr	0.59	18	0.64	18	0.40	14	0.47	14	...	...	...	...
Final setting, Gillmore, hr	0.92	17	0.89	16	0.80	16	0.85	16	...	...	...	...
Expansion, per cent	0.04	26	0.03	23	0.04	55	0.03	34	0.008	21	0.01	38
Air entrainment, per cent	1.4	20	1.3	15	1.1	14	1.9	12	1.2	12	1.4	12
Water-air entrainment, per cent	3.6	5.1	3.4	4.9	9.5	5.1	4.9	8.1	2.6	3.8	3.3	4.9
Compressive strength, 3 days, psi	221	9.5	276	9.7	250	9.0	237	11	163	8.6	231	10
Compressive strength, 7 days, psi	305	8.1	234	6.4	304	7.2	280	7.8	206	7.7	313	8.6
Water compressive strength, per cent	1.4	2.8	1.3	2.7	1.2	2.5	1.2	2.5	1.3	2.8	1.5	3.1
Fineness, *B*, sq cm per g	105	3.2	90	2.6	177	4.6	107	3.5	112	3.2	304	6.7
SiO_2, per cent	0.16	0.76	0.21	0.95	0.18	0.84	0.20	0.91	0.19	0.91	0.24	1.1
Al_2O_3, per cent	0.36	5.3	0.33	6.8	0.97	6.2	0.29	4.9	0.35	6.5	0.45	5.6
Fe_2O_3, per cent	0.20	8.1	0.24	8.3	0.22	7.9	0.21	7.5	0.12	2.5	0.10	4.1
CaO, per cent	0.48	0.76	0.43	0.69	0.39	0.61	0.43	0.66	0.87	1.4	0.70	1.3
MgO, per cent	0.21	7.3	0.28	7.4	0.15	8.3	0.14	10	0.25	11	0.24	6.2
SO_3, per cent	0.11	6.3	0.12	4.3	0.10	5.3	0.09	5.2	0.14	7.6	0.13	5.6
Ignition loss, per cent	0.12	16	0.10	13	0.14	12	0.16	21	0.22	16	0.37	14
Residue, per cent	0.15	94	0.08	32	0.10	65	0.17	96	0.08	45	0.10	20
Na_2O, per cent	0.03	8.7	0.03	11	0.03	15	0.03	19	0.03	14	0.03	36
K_2O, per cent	0.04	5.1	0.07	7.5	0.05	6.4	0.04	5.1	0.04	8.2	0.03	15
Mn_2O_3, per cent	...	...	...	...	...	...	...	...	0.07	82	0.09	15
Sulfide sulfur, per cent	...	...	...	...	...	...	...	...	0.01	71	0.07	12
False set, initial mm	...	...	...	...	...	...	...	...	...	...	...	...
False set, final mm	...	...	...	...	...	...	...	...	...	...	...	...
Tensile strength, 3 days, psi	...	...	...	...	32	9.1	29	9.1	...	...	...	...
Tensile strength, 7 days, psi	...	...	...	...	36	8.6	29	7.4	...	...	...	...
Water tensile strength, per cent	...	...	...	...	0.08	0.72	0.19	1.8	...	...	...	...
Vicat setting, hr	...	...	...	...	...	...	...	...	0.37	14	71	24
Fineness, Wagner, sq cm per g	...	...	...	...	...	...	...	...	...	...	...	...
Retained on No. 325 sieve, g	...	...	...	...	...	...	...	...	...	...	...	...
Sulfate resistance expansion, 14 days, per cent	...	...	...	...	...	...	...	...	...	...	...	...
Sulfate resistance expansion, 28 days, per cent	...	...	...	...	...	...	...	...	...	...	...	...
Heat of hydration, 7 days cal per g	...	...	...	...	...	...	...	...	...	...	...	...
Heat of hydration, 28 days, cal per g	...	...	...	...	...	...	...	...	...	...	...	...

S.D. Standard deviation.
C.V. Coefficient of variation, per cent.

Sample	No. 7		No. 8		No. 9		No. 10		No. 11		No. 12	
	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.	S.D.	C.V.
Water consistency, per cent	1.1	3.9	0.59	2.4	0.44	1.8	0.94	3.7	0.58	2.3	0.58	2.2
Initial setting, Gillmore, hr	0.39	14	0.44	15	0.51	14	0.85	18	0.46	14	0.45	15
Final setting, Gillmore, hr	0.81	17	0.87	17	0.82	14	1.3	17	0.69	13	0.64	13
Expansion, per cent	0.03	115	0.04	23	0.01	24	0.01	30	0.12	132	0.05	87
Air entrainment, per cent	1.4	20	1.2	6.2	1.3	12	1.4	15	1.5	9.1	1.3	13
Water-air entrainment, per cent	2.5	3.6	2.9	5.0	2.9	4.3	3.0	4.4	2.8	4.8	2.1	3.0
Compressive strength, 3 days, psi	240	9.0	179	10	156	8.4	193	8.4	209	9.1	254	9.2
Compressive strength, 7 days, psi	333	6.9	294	12	248	8.9	293	8.7	293	7.9	320	7.4
Water, compressive strength, per cent	1.2	2.5	1.3	2.9	1.3	2.8	1.4	3.0	1.1	2.4	1.1	2.3
Fineness, *B*, sq cm per g	279	6.1	90	2.9	121	3.6	123	4.2	135	4.3	109	2.8
SiO_2, per cent	0.21	1.0	0.22	0.99	0.36	1.7	0.34	1.5	0.34	1.6	0.25	1.2
Al_2O_3, per cent	0.20	3.6	0.30	5.9	0.29	5.1	0.36	9.7	0.22	3.7	0.24	4.0
Fe_2O_3, per cent	0.09	3.6	0.08	3.3	0.11	2.5	0.15	3.5	0.10	3.7	0.09	3.4
CaO, per cent	0.34	0.52	0.32	0.50	0.42	0.67	0.42	0.64	0.45	0.69	0.41	0.64
MgO, per cent	0.15	7.8	0.12	3.4	0.27	23	0.20	26	0.18	13	0.21	11
SO_3, per cent	0.13	4.4	0.11	5.1	0.09	4.5	0.09	5.3	0.06	3.4	0.08	4.5
Ignition loss, per cent	0.15	12	0.10	12	0.14	13	0.10	14	0.11	13	0.14	11
Residue, per cent	0.06	30	0.06	37	0.08	55	0.07	54	0.06	46	0.05	40
Na_2O, per cent	0.04	30	0.04	13	0.03	20	0.07	13	0.03	19	0.03	16
K_2O, per cent	0.04	5.2	0.08	7.5	0.04	4.7	0.05	24	0.05	5.8	0.04	5.8
Mn_2O_3, per cent	...	...	...	...	...	...	...	...	...	...	...	...
Sulfide sulfur, per cent	...	...	...	...	...	...	...	...	...	...	...	...
False set, initial mm	...	...	...	...	...	...	...	...	5.6	18	6.6	20
False set, final mm	...	...	...	...	...	...	...	...	5.4	27	9.0	48
Tensile strength, 3 days, psi	...	...	...	...	...	...	...	...	...	...	...	...
Tensile strength, 7 days, psi	...	...	...	...	...	...	...	...	...	...	...	...
Water tensile strength, per cent	...	...	...	...	...	...	...	...	...	...	...	...
Vicat setting, hr	...	...	...	...	...	...	...	...	...	...	...	...
Fineness, Wagner, sq cm per g	119	4.7	95	5.5	...	...	...	...	...	...	...	...
Retained on No. 325 sieve, g	0.02	131	0.02	18	...	...	...	...	...	...	...	...
Sulfate resistance expansion, 14 days, per cent	...	...	...	...	0.002	38	0.001	49	...	...	...	...
Sulfate resistance expansion, 28 days, per cent	...	...	...	...	0.022	123	0.011	122	...	...	...	...
Heat of hydration, 7 days, cal per g	...	...	...	...	4.2	6.1	5.1	7.9	...	...	...	...
Heat of hydration, 28 days, cal per g	...	...	...	...	7.8	9.5	8.9	11	...	...	...	...

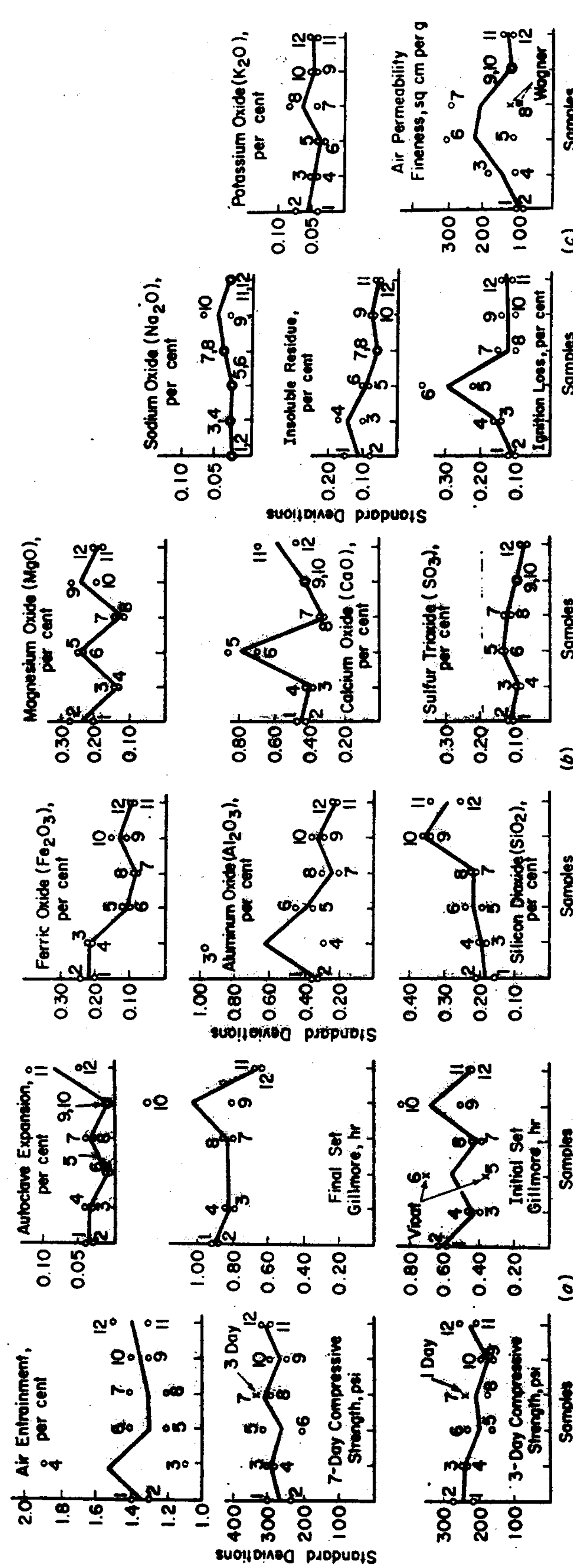

FIG. 1.—Average Standard Deviations for the Six Successive Pairs of Cement Samples.

The participating laboratories were requested to make the physical and chemical tests as required by detailed instructions furnished with each pair of samples and to report the results to the National Bureau of Standards within 1 month after receiving the samples. With the instructions for the first pair of samples it was also requested that the same operator and the same analyst of each laboratory also make all subsequent test determinations.

As the reports were received and the tabulated results scanned, those laboratories whose result appeared to contain an error were asked to confirm the questionable value. In most instances, suspicion of the presence of an error was confirmed in time to permit inclusion of the corrected value in the program. At the NBS, the results were tabulated, the data placed on IBM cards, and the averages, standard deviations between laboratories, and coefficients of variation computed by a digital computer. Scatter diagrams as proposed by Youden[3] were plotted for each set of test results. The scatter diagrams, averages, standard deviations, and coefficients of variation were furnished each of the participating laboratories at the time the next pair of samples was distributed. This afforded the personnel of the participating laboratories an opportunity to study the results of the previous tests and make procedural corrections before making tests on the new samples.

RESULTS OF TESTS

The standard deviations between laboratories and corresponding coefficients of variation for data reported by all laboratories are presented in Table IV. The values for standard deviation are also presented in the graphs shown in Figs. 1(*a*), (*b*), and (*c*). It may be noted that considerable differences exist in the variability among laboratories for the different tests, for the different samples in pairs, and for the different pairs. For most tests there was no definite evidence of improvement of results as the program proceeded. The use of referee chemical methods did not result in better concordance than was obtained with the use of optional chemical methods. The average values for the various tests on the different samples are presented in Appendix II.

Scatter diagrams were prepared in accordance with the method proposed by Youden[3] in which the value for a property of one sample is plotted on the horizontal axis and the corresponding value for the other sample on the vertical axis. Thus the results for one property for two samples from each laboratory are represented by a point on the diagram. Each diagram is divided into quadrants by a vertical and a horizontal line to indicate the medians. That is, there are as many laboratories to the left of the vertical line as to the right and there are as many laboratories above the horizontal line as below.

Figure 2 shows the distribution of results for the percentage loss on ignition of cement samples Nos. 11 and 12. The pattern formed by the points is fairly circular and the crossed center lines are the median values for the samples. Only two of the laboratories had ignition loss results for these two cements that deviated considerably from the median values, and there was a fairly uniform distribution of the points. Under ideal circumstances, there should be nearly an equal number of points in each of the four quadrants and the pattern of points

[3] W. J. Youden, "A New Graphic Method for Statistical Treatment and Evaluation of Interlaboratory Tests," "Application of the Graphic Method to Other Interlaboratory Testing and Specifications." These papers presented at the Sixty-second Annual Meeting of the Society, June 21–26, 1959, have been combined under the title "Statistical Aspects of the Cement Testing Program," see p. 1120, this publication.

should be roughly circular, as shown in Fig. 2.

When this method of presentation is used, it is an easy matter for a participating laboratory to determine the relationship of its results to the median values and to the results obtained by the other laboratories.

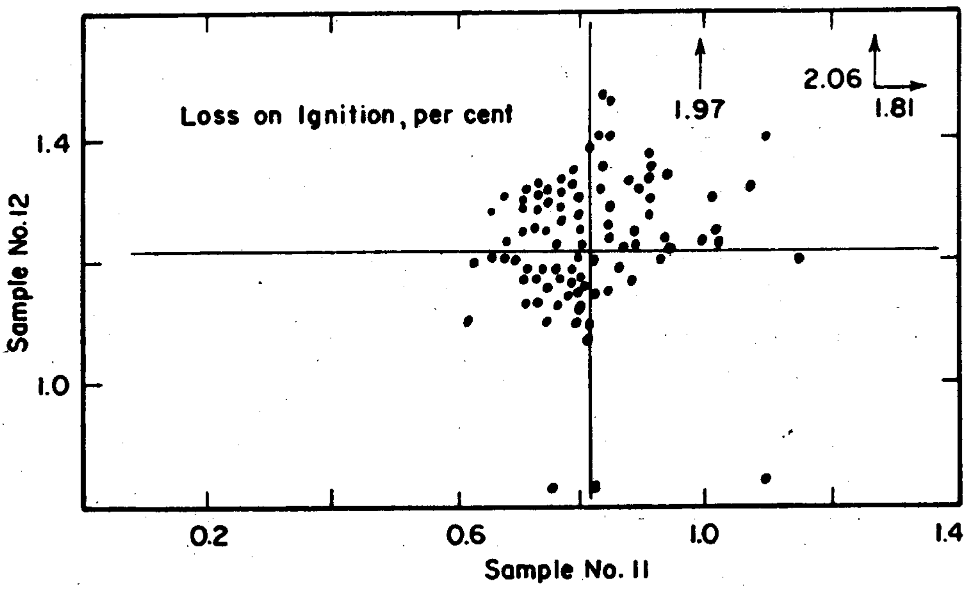

FIG. 2.—Scatter Diagram for Percentage Loss on Ignition for Samples Nos. 11 and 12.

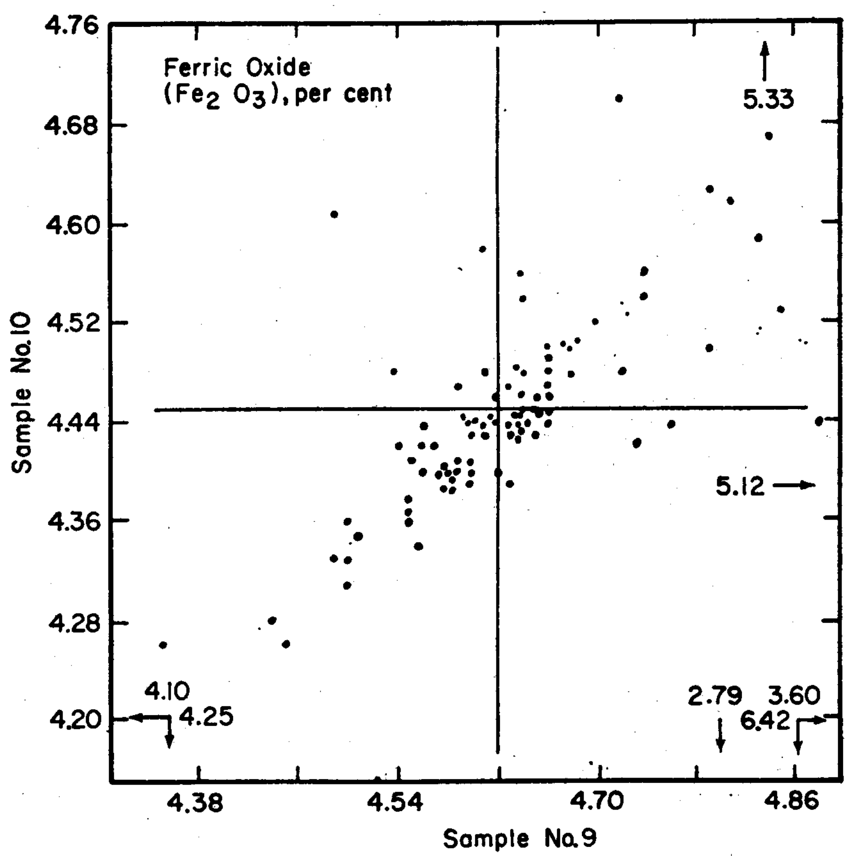

FIG. 3.—Scatter Diagram for Percentage Ferric Oxide (Fe_2O_3) for Samples Nos. 9 and 10.

If the scatter of points is oval, as indicated in Fig. 3, with the long axis of the oval at approximately 45 deg, passing through the lower left quadrant and the upper right quadrant, there is evidence of a strong correlation between the test results obtained in any one laboratory. The majority of the scatter diagrams tended to be oval along the 45 deg axis. This indicates that many of the laboratories have a tendency to obtain high values on both samples if high val-

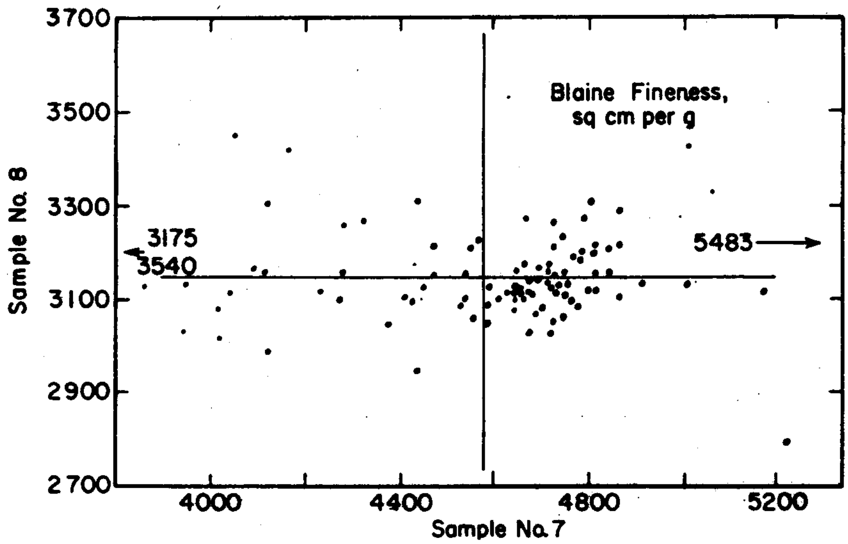

FIG. 4.—Scatter Diagram for Air-Permeability Fineness for Samples Nos. 7 and 8.

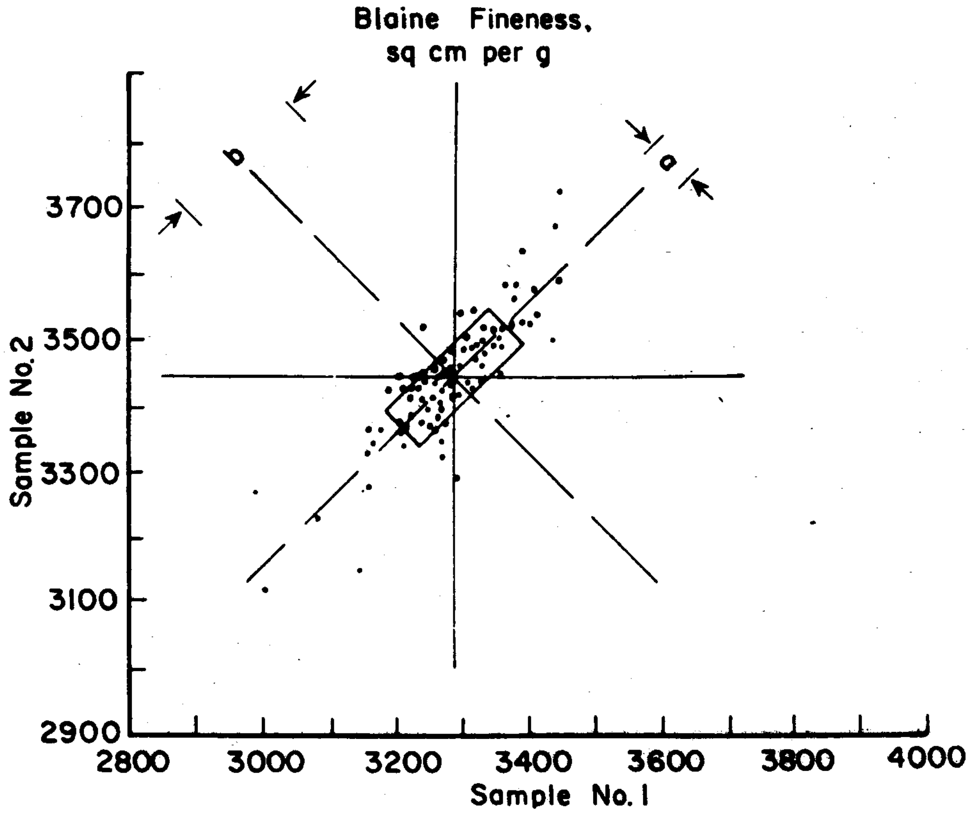

FIG. 5.—Scatter Diagram for Air-Permeability Fineness for Samples Nos. 1 and 2. The significance of parameters "a" and "b" is explained in the text.

ues are obtained on one. In this particular case, there should be 25 points in each quadrant. However, the lower left and upper right quadrants together account for about 80 points, which is evidence of numerous laboratory biases. Among the possible sources of such errors may be the ambiguity or misinterpretation of the test methods. With a chemical analysis such as this, errors in standardization of solutions or possibly constant errors in weighing could contribute to the bias. With uniform samples, a broader oval normally indicates a greater imprecision on the part of the operator or apparatus. Laboratories whose results are consistently represented by points fairly far out in the upper right or lower left quadrant are believed to have a systematic error in the test procedure or apparatus.

TABLE V.—MEASUREMENT OF STANDARD DEVIATION OF TEST METHOD AND LABORATORY BIAS FROM SCATTER DIAGRAMS.
Averages for All Twelve Samples

	a	*b/a*
Physical Tests		
Normal consistency, per cent of H_2O	1.0	1.47
Time of setting, initial, hr	0.72	2.03
Time of setting, final, hr	0.80	2.71
Autoclave expansion, per cent	0.023	1.56
Air entrainment, per cent	1.26	2.50
Compressive strength, 3 days, psi	255	1.86
Compressive strength, 7 days, psi	314	2.12
Fineness, sq cm per g	162	1.85
Chemical Tests		
Silicon dioxide (SiO_2), per cent	0.17	2.17
Aluminum oxide (Al_2O_3), per cent	0.26	2.24
Ferric oxide (Fe_2O_3), per cent	0.07	2.33
Calcium oxide (CaO), per cent	0.25	2.83
Magnesium oxide (MgO), per cent	0.14	2.14
Sulfur trioxide (SO_3), per cent	0.07	1.97
Ignition loss, per cent	0.18	1.58
Insoluble residue, per cent	0.06	2.35
Sodium oxide (Na_2O), per cent	0.02	2.96
Potassium oxide (K_2O), per cent	0.05	2.10

Figure 4 shows a pattern which indicates no laboratory bias, but a greater spread for one sample than for the other. This results in a pattern which is oval in form with the long axis either vertical or horizontal, depending on which sample shows the greater scatter for the test results. Sample No. 7 was a high early-strength cement, and it is quite apparent that the laboratories had difficulty determining the surface areas of this much finer material. Few instances were found having this type of scatter diagram.

An evaluation of both the laboratory bias and the test method precision was obtained from the parameters of rectangles constructed as shown in Fig. 5. The sides of the rectangles were formed by pairs of lines drawn parallel to and equidistant from the intersecting 45 deg lines. For each direction, the parallel lines were spaced so that approximately 68 per cent of the points fell between them. This 68 per cent should be within plus or minus one standard deviation in a normal distribution. The magnitude of a is an indication of the imprecision of the test. A ratio of b/a greater than 1 is an indication of a laboratory bias which is significantly greater than the random error of the test. Obviously, it would be desirable to have the ratio b/a as close to 1 as possible, and it would be desirable to have a as small as possible. This type of graph indicates where to look for possible improvement of test results.

The values for a and b/a obtained in this program are presented in Table V. The values are averages for all 12 samples. Tests for final time of set and air entrainment have the largest b/a ratio or laboratory bias in the physical tests, whereas Na_2O and CaO determinations have the largest b/a ratio in the chemical tests. The chemical tests had an average b/a ratio of 2.27, as compared to an average of 2.01 for the physical tests. A lack of precision in the test method itself, that is, a large value for a, would result in a lower b/a ratio.

After the tests on the first three pairs of samples had been completed, the average standard deviation for each of the tests was determined. Figure 6 is the scatter diagram for air entrainment for samples Nos. 7 and 8 with squares drawn using the standard deviation thus determined. The sides of the dotted inner square are plus and minus one times this standard deviation from the center. The sides of the outer square are at 1.95 times the standard deviation from the center, and this larger square should contain 90 per cent of all the points. These scatter diagrams were distributed to all the laboratories, giving them an opportunity to evaluate their own work in relation to that done by all the laboratories and to re-examine and improve their own procedures where need for improvement was indicated.

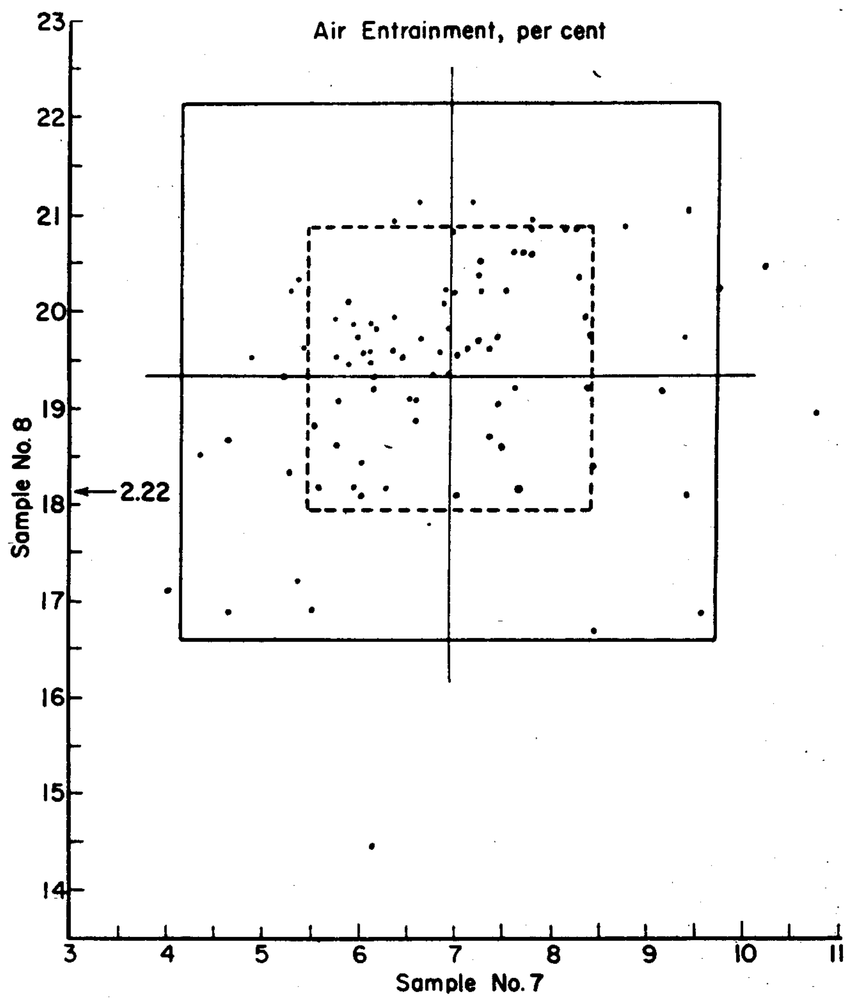

FIG. 6.—Scatter Diagram for Percentage Air Entrained in the 1:4 Mortars Reported for Samples Nos. 7 and 8.

The significance of the squares is explained in the text.

The graphic methods just discussed would enable a laboratory to determine its proficiency with respect to any one of the tests. In order to develop an evaluation of a laboratory's performance on all tests, the laboratories were rated on the basis of the averages and standard deviations obtained. A schematic diagram of the method used for rating the values reported by the various laboratories is shown in Fig. 7. The center block represents values within plus and minus one standard deviation of the mean. These values were assigned a

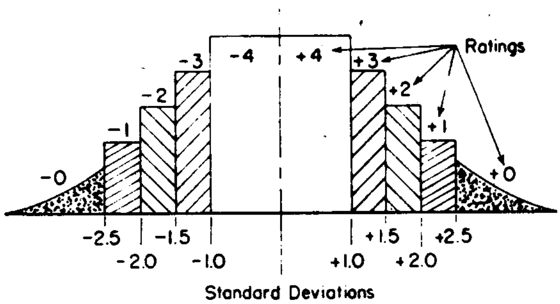

FIG. 7.—System Used to Assign Ratings to Laboratories on the Basis of Values Reported by All the Participants.

rating of 4. Deviations from the mean of 1.0 to 1.5 times the standard deviations were assigned the rating 3. Between 1.5 and 2.0 standard deviations, the rating was 2, and from 2.0 to 2.5 standard deviations it was 1. Any value greater than 2.5 times the standard deviation was rated 0. Values to the right of center or above average were designated plus, and those to the left of center or below average were designated minus.

Table VI demonstrates the application of this rating system. Laboratory A, in determining normal consistency of neat pastes, and laboratory B, in determining SO_3 obtained values close to the averages of all laboratories for the tests in question. On the other hand, the results obtained by laboratory C, in determining normal consistency of neat pastes, indicates the presence of a consistent bias with respect to the average of all laboratories. Laboratory D, in determining autoclave expansion, had a very low rating at the start, but rapidly moved closer to the laboratory average. This laboratory had a tendency to obtain low results for this test. The ratings for laboratory E, determining 7-day compressive strength, show a great deal of variability. Laboratory F, determining normal consistency of neat pastes, was slightly below the general average.

These data indicate the value of a continuous reference sample program in helping a laboratory to continually evaluate its results.

Average values for ratings (without regard to signs) were computed for all the physical and all the chemical tests for each laboratory for each sample tested.

TABLE VI.—EXAMPLES OF RATINGS OBTAINED BY SIX OF THE LABORATORIES ON 12 SAMPLES.

Samples		No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11	No. 12	Average[c]
Laboratory	Test													
A	a[a]	+4	−4	+4	+4	+4	+4	−4	+4	+4	−4	−4	−4	4.00
B	b	4[b]	+4	+4	+4	+4	+4	−4	−4	4[b]	+4	−4	−4	4.00
C	a	+1	+3	+3	+0	+0	+2	+3	+0	+3	+2	+0	+0	1.42
D	c	−0	−0	−4	−4	−3	4[b]	−4	4[b]	−3	+1	−4	−4	2.91
E	d	−4	+4	−1	−2	−4	−0	+2	+4	+2	+1	−3	−2	2.42
F	a	+2	+2	−4	−3	...	...	−3	+4	−4	−4	−1	+3	3.00

[a] Tests *a* = Normal consistency.
b = Sulfur trioxide (SO_3).
c = Autoclave expansion.
d = Compressive strength, 7 days.
[b] Identical with average for the test for this sample.
[c] Average computed without regard to signs.

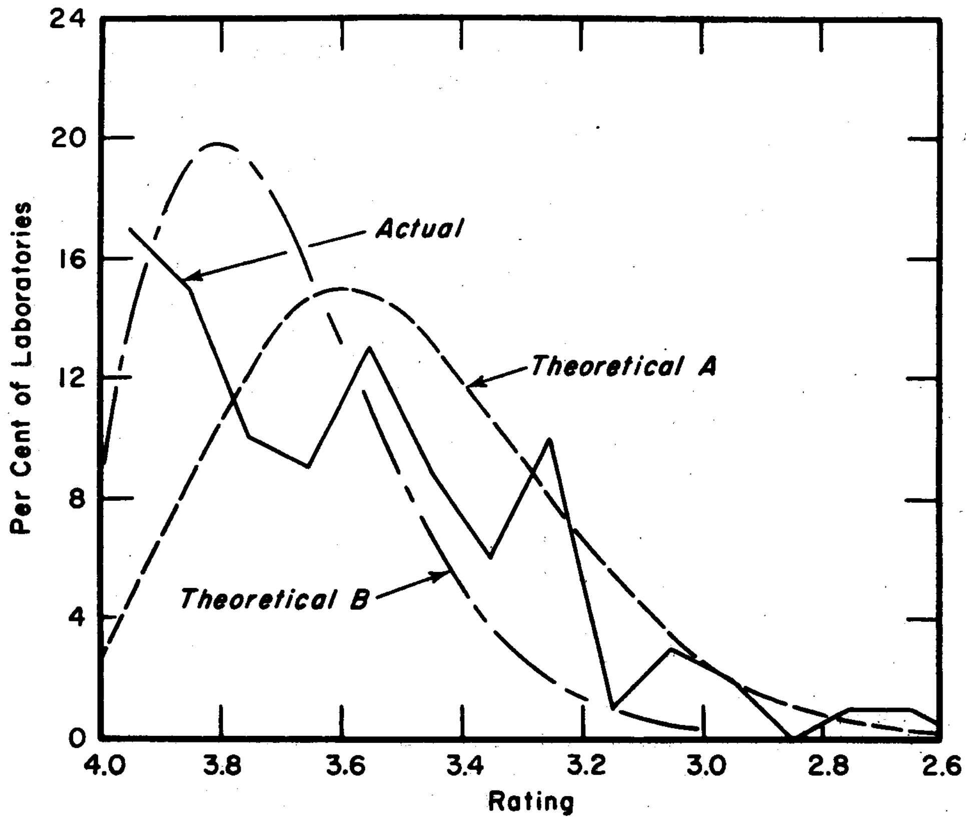

FIG. 8.—Frequency Distribution of Average Laboratory Ratings for All Physical Tests Reported for Sample No. 6.

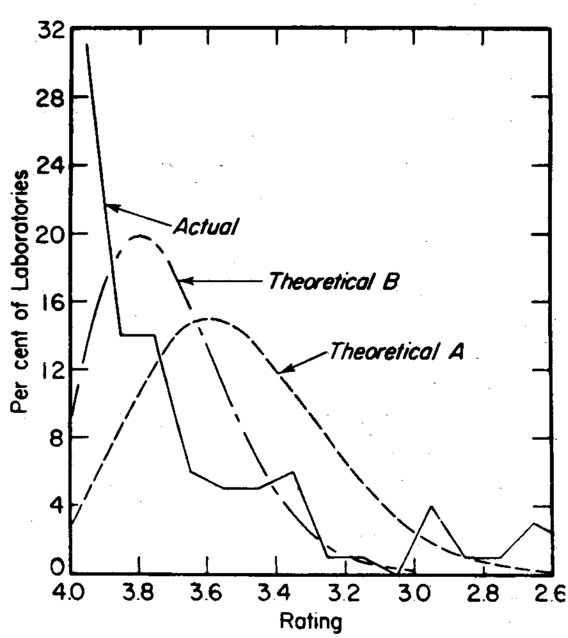

FIG. 9.—Frequency Distribution of Average Laboratory Ratings for All Chemical Tests Reported for Sample No. 6.

The distribution of the ratings for all 10 physical and 10 chemical tests made on sample No. 6, together with two theoretical curves, are presented in Figs. 8 and 9 respectively. These were typical of the distribution obtained on all 12 samples for each test method.

Theoretical curve *A* was based on averages of ten scores where individual scores were assigned as shown in Fig. 7. The standard deviations used are those shown in Table IV which were computed by using all results except obvious blunders. One possible interpretation of the disagreement between curve *A* and the observed frequency distribution of the laboratory scores is in the assumption that all but a small number of laboratories are in better agreement than is reflected in these standard deviations and that the latter are inflated through the inclusion of the excessively discrepant laboratories. This would imply, for example, that some results which fell in this range and were awarded a score of 4 should have had a lower score. Consequently, the proportion of high scores actually awarded was greater than predicted by curve *A*. Curve *B* was derived by recomputing the probabilities for each score, making the arbitrary assumption that the actual standard deviation was 80 per cent of that shown in Table IV. The elimination of the results from about six laboratories with lowest ratings would bring this reduction in the standard deviation. The theoretical distribution of the average scores was calculated and adjusted to 94 laboratories instead of 100, as used for curve *A*. Curve *B* is a better approximation to the observed distribution of average scores and, therefore, tends to support the smaller standard deviations as better estimates of the actual performance of over 90 per cent of the laboratories.

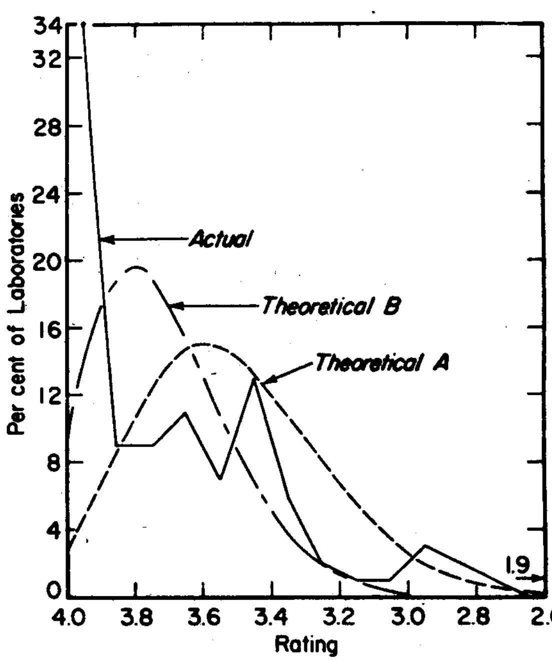

FIG. 10.—Frequency Distribution of Average Laboratory Ratings for Air Permeability Fineness Values Reported for All Twelve Samples.

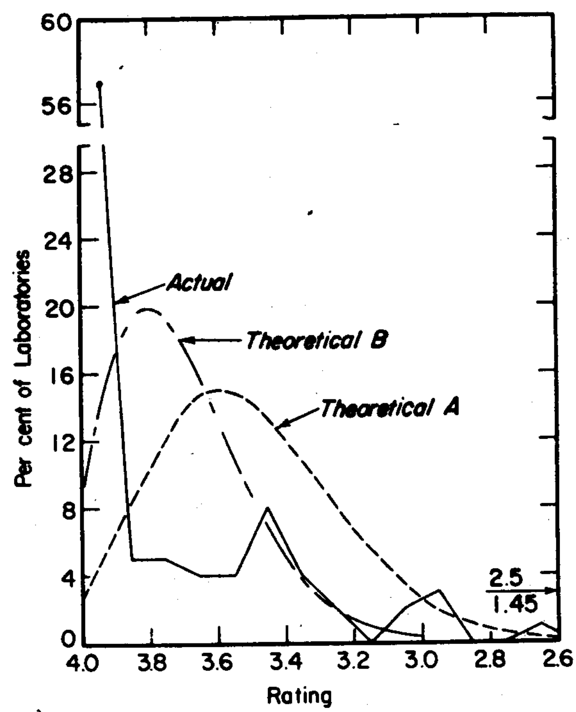

FIG. 11.—Frequency Distribution of Average Laboratory Ratings for Percentage Sulfur Trioxide (SO_3) Values Reported for All Twelve Samples.

In drawing the above conclusions from the frequency distribution of the scores, it should be borne in mind that both theoretical curves are based on the assumption that the scores obtained by any given laboratory for different tests or different samples are statistically independent. An examination of the individual scores reveals, however, that such scores tend to be correlated in a number of cases over the entire sampled population of laboratories. This would indicate that a laboratory tends to maintain its position (as indicated by the scoring system) for different samples on the same test and even, in some cases, for different tests. The disagreement between the theoretical curve *A* and the observed frequency distribution of scores must then be ascribed, in part at least, to those correlations, and this possibility is supported by the still unsatisfactory agreement of the observed distribution of scores with curve *B*.

But whether the distortion of the frequency distribution of the scores is ascribed primarily to the inflation of the standard deviations through inclusion of a few discrepant laboratories or to the correlation between scores obtained by the same laboratory, in either case the samples and for one chemical test for all 12 samples, together with the two theoretical frequency distribution curves. The data show the presence of recognizable differences in the performance of the laboratories.

TABLE VII.—FREQUENCY DISTRIBUTIONS OF AVERAGE RATINGS OF ALL LABORATORIES REPORTING PHYSICAL TEST RESULTS.

Tests	a	b	c	d	e	f	g	h	i	j	Total	Average, per cent
Ratings												
4.00 to 3.91	18	17	20	35	27	33	17	21	30	34	252	24.53
3.90 to 3.81	11	17	16	9	9	10	17	12	15	9	125	12.17
3.80 to 3.71	20	6	15	9	9	15	8	13	12	9	116	11.29
3.70 to 3.61	10	9	7	12	12	7	13	9	13	11	103	10.02
3.60 to 3.51	7	10	10	7	7	9	9	5	2	7	73	7.10
3.50 to 3.41	16	10	4	11	13	3	12	14	8	13	104	10.12
3.40 to 3.31	7	5	5	5	7	6	4	7	4	6	56	5.45
3.30 to 3.21	5	12	3	4	3	3	4	8	5	2	49	4.77
3.20 to 3.11	2	6	8	4	3	5	3	4	2	1	38	3.70
3.10 to 3.01	1	2	4	2	2	2	4	1	1	1	20	1.94
3.0 to 2.91	1	1	2	2	3	1	7	2	2	3	24	2.33
2.90 to 2.81	2	1	...	...	2	2	2	2	1	2	14	1.36
2.80 to 2.71	...	2	1	1	...	2	...	1	2	1	10	0.97
2.70 to 2.61	...	...	...	...	2	2	1	...	1	...	6	0.58
2.60 to 2.51	...	...	1	...	1	...	...	1	1	...	4	0.38
2.50 to 2.41	1	1	1	...	...	1	...	3	...	...	7	0.69
2.40 to 2.31	...	1	2	...	1	...	2	...	1	...	7	0.68
2.30 to 2.21	...	...	1	...	...	...	...	...	...	...	1	0.097
2.20 to 2.11	...	1	...	...	...	...	...	...	2	...	3	0.294
2.10 to 2.01	...	1	...	...	...	...	...	...	...	...	1	0.097
2.00 to 1.91	...	...	1	...	...	1	...	...	...	...	2	0.194
1.90 to 1.81	...	...	1	...	...	...	...	...	...	1	2	0.194
1.80 to 1.71	...	...	...	...	...	...	...	...	...	...	...	...
1.70 to 1.61	...	1	...	...	...	...	...	...	...	...	1	0.097
1.60 to 1.51	...	...	...	...	...	1	...	...	...	...	1	0.097
1.50 to 1.41	1	...	...	...	1	...	...	...	...	...	2	0.194
1.40 to 1.31	...	...	...	1	...	...	...	...	...	...	1	0.097
1.30 to 1.21	...	...	1	...	...	...	...	...	...	...	1	0.097
1.20 to 1.11	...	...	...	...	1	...	...	...	...	...	1	0.097
1.10 to 1.01	...	...	...	1	...	...	...	...	...	...	1	0.097
1.00 to 0.91	1	...	...	...	...	...	...	...	...	...	1	0.097
Total	103	103	103	103	103	103	103	103	103	100	1027	99.82

[a] = Normal consistency test.
[b] = Initial setting, Gillmore test.
[c] = Final setting, Gillmore test.
[d] = Autoclave expansion test.
[e] = Air entrainment test.
[f] = Water for air entrainment test.
[g] = Compressive strength, 3 days, test.
[h] = Compressive strength, 7 days, test.
[i] = Water for compressive strength test.
[j] = Air permeability fineness test.

In Figs. 10 and 11 are presented the frequency distributions of the laboratory ratings for one physical test for all 12 patterns indicated in both cases are similar to those indicated in Figs. 8 and 9.

These two curves also are typical of the 10 chemical tests and the 10 physical tests performed on each of the 12 samples.

The frequency distributions of the ratings of the laboratories for each of the 10 physical tests for the 12 samples are presented in Table VII, and, for the 10 chemical tests, in Table VIII. The majority of the laboratories had satisfactory ratings, although many had low or very low ratings. As pointed out by Youden,[3] the low ratings of a few laboratories significantly affected the over-all ratings of all laboratories.

Further studies of the frequency distribution of scores are illustrated in Fig. 12. In this figure the frequency distribution of the laboratory ratings for all physical and chemical tests for all samples are presented with the two theoretical curves. It may be noted that more laboratories had ratings of 4.0 to 3.9 for both chemical and physical tests than would be predicted from the theoretical curves *A* or *B*. The number of laboratories with ratings of 3.8 to 3.5 was less and the number of laboratories with very low ratings was more than would be predicted from the theoretical curves. Figures 8 to 11

TABLE VIII.—FREQUENCY DISTRIBUTIONS OF AVERAGE RATINGS OF ALL LABORATORIES REPORTING CHEMICAL TEST RESULTS.

Tests	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	SO_3	Loss on Ignition	Residue	Na_2O	K_2O	Total	Average, per cent
Ratings												
4.00 to 3.91	43	37	61	45	41	57	48	43	27	25	427	45.32
3.90 to 3.81	4	15	14	16	8	5	7	15	5	10	99	10.50
3.80 to 3.71	9	9	...	6	8	5	4	5	7	5	56	5.94
3.70 to 3.61	9	3	5	5	5	4	10	6	6	1	58	6.15
3.60 to 3.51	5	7	5	4	5	4	5	6	6	3	50	5.30
3.50 to 3.41	10	7	4	7	7	8	7	5	3	7	65	6.90
3.40 to 3.31	4	4	2	5	2	4	2	3	2	4	34	3.60
3.30 to 3.21	1	3	...	1	6	2	1	4	...	3	17	1.80
3.20 to 3.11	1	5	3	3	2	...	3	...	3	1	21	2.22
3.10 to 3.01	1	2	...	...	3	2	...	2	1	2	13	1.38
3.00 to 2.91	6	2	1	1	4	3	4	3	2	2	28	2.97
2.90 to 2.81	1	...	1	1	2	...	1	1	...	1	8	0.84
2.80 to 2.71	...	...	1	...	1	...	1	1	1	1	6	0.63
2.70 to 2.61	1	...	...	1	...	1	...	1	...	...	4	0.42
2.60 to 2.51	...	...	...	...	1	...	1	1	...	...	3	0.31
2.50 to 2.41	1	...	1	2	1	3	3	3	2	...	16	1.69
2.40 to 2.31	...	...	...	...	1	...	1	...	1	...	3	0.31
2.30 to 2.21	1	...	...	...	1	...	...	1	...	...	3	0.31
2.20 to 2.11	...	...	...	...	1	...	...	1	1	...	3	0.31
2.10 to 2.01	...	...	1	...	...	...	...	...	...	1	2	0.21
2.00 to 1.91	1	...	...	1	...	...	1	...	...	1	4	0.42
1.90 to 1.81	1	1	...	...	...	...	...	...	...	...	2	0.21
1.80 to 1.71	...	...	1	...	1	...	...	...	...	...	2	0.21
1.70 to 1.61	1	...	...	...	...	...	1	...	...	...	2	0.21
1.60 to 1.51	...	2	...	...	...	...	1	...	...	...	3	0.31
1.50 to 1.41	1	...	...	...	...	1	...	...	...	...	2	0.21
1.40 to 1.31	...	...	...	1	1	1	...	...	...	...	3	0.31
1.10 to 1.01	...	...	1	...	...	...	...	...	...	...	1	0.10
1.00 to 0.91	...	...	1	...	...	...	...	...	...	...	1	0.10
0.90 to 0.81	...	...	2	...	...	...	...	...	...	...	2	0.21
0.80 to 0.71	...	...	...	...	1	...	...	...	...	...	1	0.10
0.70 to 0.61	...	...	...	1	...	...	...	...	...	...	1	0.10
0.60 to 0.51	...	...	...	...	...	...	...	...	...	...	...	...
0.50 to 0.41	...	...	...	1	...	...	...	...	...	...	1	0.10
0.40 to 0.31	...	...	...	...	...	...	1	...	...	...	1	0.10
Total	101	101	101	101	101	101	101	101	67	67	942	99.80

similarly indicate departures from the theoretical curves.

These tables and curves show that an improvement in the ratings of a relatively small proportion of the laboratories would result in a considerable improvement in the standard deviations of all tests.

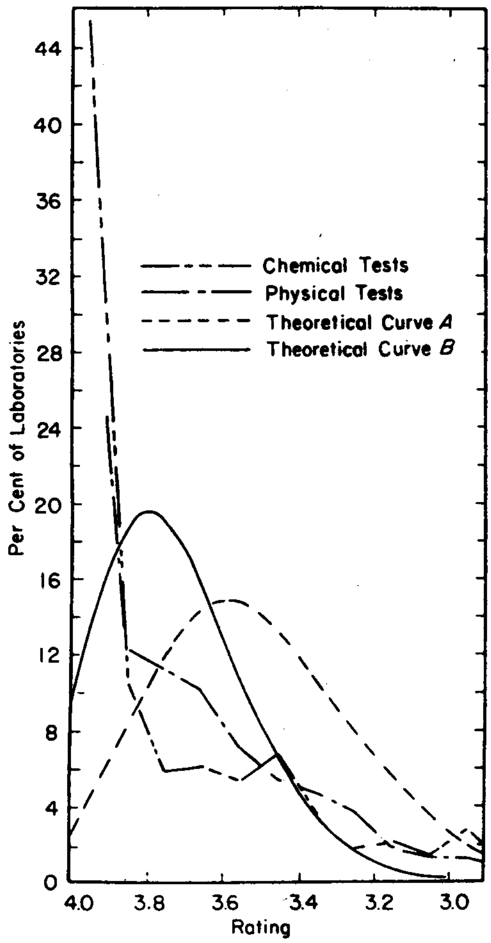

FIG. 12.—Frequency Distribution of Average Laboratory Ratings for All Physical Tests and All Chemical Tests Reported for All Twelve Samples.

On the basis of the data obtained in this program, it might be suggested that any laboratory that consistently has a rating of less than 3.5 by this rating system might well look into the possible causes for disagreement with the majority of laboratories.

As indicated earlier, all data were included in the tabulations and calculations except the obvious errors. Evidence has been presented that some of the data included were subject to extreme systematic errors and should have been excluded. It therefore appeared desirable to use another method for rating the laboratories.

Six large cement manufacturing companies were asked to advise NBS as to the tolerance limits used by them for each of the various physical and chemical tests in their interlaboratory test programs. These six tolerances, averaged for each of the tests, were used to re-evaluate the performance of all the participating laboratories. The tolerances were applied to the averages for the values reported by all the laboratories for each of the tests.

Table IX lists for each of the physical and chemical tests of the 12 samples, the average standard deviations, the averages of the tolerances suggested by the 6 laboratories, the percentage of laboratories passing these tolerances, and also the values for $a/2$ (see Table V). It may be noted that the values for the average standard deviations developed in this program and the average tolerances of the 6 cement laboratories were very similar in the case of both the physical and chemical tests. The percentage of laboratories passing the tolerances was generally higher in chemical than in physical tests.

The values for a presented in Table V represented the spread of 68 per cent of the results closest to the 45-deg line. It has previously been shown that these values are a measure of the random errors of the test methods. With normal distribution, the results of this same number of laboratories, 68 per cent, should be within plus or minus one standard deviation of the mean, and half of this value, or $a/2$, should correspond to the standard deviation of the test method with the effect of laboratory bias eliminated. Values of $a/2$ presented in Table IX for the various tests may be compared with the average standard deviation values obtained in this test program for the physical and chemical tests of the various cements and to the average tolerances of the 6 laboratories. It may be noted that the values for the average standard deviations for the different tests were, in most instances, 2 to 3 times the corresponding values for $a/2$, the standard deviation values attributable to the precision of the test methods.

TABLE IX.—A COMPARISON OF THE STANDARD DEVIATION VALUES ATTRIBUTABLE TO THE TEST METHODS, $a/2$,[a] THE STANDARD DEVIATION VALUES FOR THE 12 SAMPLES, AND THE TOLERANCE VALUES PERMITTED BY 6 LABORATORIES WHO REGULARLY CONDUCT INTERLABORATORY TESTS.

	$\frac{a}{2}$	Average Standard Deviation, 12 Samples	Permitted Tolerance of 6 Laboratories, average	Per Cent Laboratories Passing Tolerance of 6 Laboratories
PHYSICAL TESTS				
Initial setting, hr	0.36	0.52	±0.50	71
Final setting, hr	0.40	0.86	±0.66	62
Expansion, per cent	0.012	0.03	±0.06	95
Air entrainment, per cent	0.65	1.40	±1.25	71
Compressive strength, 3 days, psi	128	218	±10 %	76
Compressive strength, 7 days, psi	157	285	±10 %	83
Fineness, Blaine, sq cm per g	81	146	±100	71
CHEMICAL TESTS				
SiO_2, per cent	0.085	0.24	±0.21	76
Al_2O_3, per cent	0.130	0.36	±0.22	69
Fe_2O_3, per cent	0.035	0.14	±0.12	89
CaO, per cent	0.125	0.47	±0.27	70
MgO, per cent	0.070	0.20	±0.21	83
SO_3, per cent	0.035	0.11	±0.11	87
Ignition loss, per cent	0.090	0.16	±0.19	88
Insoluble residue, per cent	0.030	0.09	±0.11	91
Na_2O, per cent	0.010	0.04	±0.04	85
K_2O, per cent	0.025	0.05	±0.04	79

[a] See Table V.

ADDITIONAL TESTS

At various times during the program, the laboratories were asked to perform additional tests. Some of these were by ASTM methods, such as the tensile briquette test (C 109 – 58), Vicat time of setting (C 191 – 58), Wagner turbidimeter for fineness (C 115 – 58), heat of hydration (C 186 – 58), and false setting (C 359 – 56 T).[4] The sulfate resistance test used was one that was under study and was not an ASTM test method. The standard deviations and coefficients of variation for these tests are of about the same order of magnitude as those obtained with the other tests performed on all 12 samples (see Table IV).

Several other variations in procedure were introduced into the test program in order to determine the effect of certain variables.

With two separate pairs of samples, namely, Nos. 3 and 4 and Nos. 12 and 11, half the laboratories performed their chemical tests on cement samples that had been more thoroughly blended and

[4] These methods of test appear in the 1958 Book of ASTM Standards, Part 4.

then packaged in sealed vials, while the other half of the laboratories performed their chemical tests on portions of cement taken from the 12-lb samples used for all other tests.

With one pair of samples, Nos. 7 and 8, half of the laboratories were asked to make strength cubes using regular tap water and to store the cubes in the usual manner until time for breaking, while the other half used distilled water in making the cubes, after which the cubes were stored in a moist plastic bag in the moist cabinet until time for breaking.

Another of the variables in this program was the storage of the samples previous to distribution. It would be desirable for a laboratory, for example, to blend and package the samples during the winter season when the employees are not so busy. Half of lots 3 and 4 were therefore stored in sealed metal drums and half in the shipping packages on a platform skid, in a store room. These lots were shipped out 10 months later as samples Nos. 12 and 11 respectively.

The variables of (1) vial *versus* bulk samples, (2) tap *versus* distilled water, and (3) skid *versus* drum storage were examined by alloting 50 of the laboratories to one choice and the other 50 laboratories to the other choice. Consequently, in looking for an effect of any of the three pairs of alternatives listed above, the averages and standard deviations of 50 laboratories using one choice were compared with the averages and standard deviations of the other 50 laboratories using the other choice. These averages, standard deviations, and coefficients of variation are presented in Appendix III. It was concluded, from a study of the data, that in almost all cases, changes in these variables did not produce differences which were statistically significant. Additional data and more precise test procedures may throw

TABLE X.—FREQUENCY DISTRIBUTION OF COMPOUND COMPOSITION VALUES COMPUTED FROM CHEMICAL ANALYSIS VALUES REPORTED FOR SAMPLE NO. 9.

C_3A, per cent	Number of Laboratories	C_3S, per cent	Number of Laboratories	C_2S, per cent	Number of Laboratories	C_4AF, per cent	Number of Laboratories
1.1 to 2.0	3	33.1 to 34.0	1	16.1 to 17.0	1	12.1 to 13.0	1
2.1 to 3.0	1	34.1 to 35.0	0	...	0	13.1 to 14.0	47
3.1 to 4.0	0	35.1 to 36.0	2	24.1 to 25.0	1	14.1 to 15.0	45
4.1 to 5.0	0	36.1 to 37.0	1	25.1 to 26.0	1	15.1 to 16.0	1
5.1 to 6.0	3	37.1 to 38.0	3	26.1 to 27.0	5	16.1 to 17.0	0
6.1 to 7.0	21	38.1 to 39.0	7	27.1 to 28.0	1	17.1 to 18.0	0
7.1 to 8.0	48	39.1 to 40.0	11	28.1 to 29.0	6	18.1 to 19.0	0
8.1 to 9.0	16	40.1 to 41.0	14	29.1 to 30.0	16	19.1 to 20.0	1
9.1 to 10.0	3	41.1 to 42.0	18	30.1 to 31.0	24		
		42.1 to 43.0	18	31.1 to 32.0	12		
		43.1 to 44.0	5	32.1 to 33.0	14		
		44.1 to 45.0	5	33.1 to 34.0	6		
		45.1 to 46.0	3	34.1 to 35.0	3		
		46.1 to 47.0	3	35.1 to 36.0	3		
		47.1 to 48.0	2	36.1 to 37.0	1		
		...					
		...					
		...					
		...					
		...					
		58.1 to 59.0	1				

further light on the effect of these variables.

PRACTICAL IMPORTANCE OF PRECISE CEMENT TESTING

Precise measurements of the physical and chemical properties of cement are desirable for both manufacturing control and acceptance testing, as well as for research on the effect of these properties of the cement on the properties of concretes made with cements of different composition. For example, the compound composition calculated from the chemical analysis is one of the factors used to distinguish between types I, II, IV, and V cements. As an illustration of the effect of chemical analysis on the calculated compound composition, Table X has been prepared. These compound-composition values were computed from the chemical analyses of sample No. 9 reported by 95 of the laboratories participating in this program. The frequency distributions of the calculated percentages of C_3A, C_3S, C_2S, and C_4AF are presented in Table X. Four of the laboratories would have considered sample No. 9 a type V cement (less than 5.0 per cent C_3A); 19 would have considered it a type I cement (more than 8 per cent C_3A); and the other laboratories would have considered it a type II cement. In view of the fact that present cement specifications permit up to 3.0 per cent SO_3 when the C_3A is greater than 8.0 per cent and only 2.5 per cent SO_3 when the C_3A is 8.0 per cent or less, 19 of the laboratories would have permitted 3.0 per cent SO_3, whereas the others would have limited the SO_3 to 2.5 per cent. If estimates were made of probable heat of hydration or sulfate resistance from chemical analyses varying as much as those reported in this program, these estimates would not be of very great value.

Similarly, the wide range in results obtained in physical and chemical testing of cement might lead to controversies between the manufacturer and purchaser and might give a false impression of the properties of the cement used in a concrete structure.

Greater precision in test methods and laboratory technique would greatly reduce the margin of safety manufacturers must employ to insure that their product will meet specifications when tested by some other laboratory.

OPERATION OF A CONTINUING STANDARD REFERENCE SAMPLE PROGRAM

When the use of standard cement samples was proposed to ASTM Committee C-1 in 1957, it was pointed out that a continuing Standard Sample program would give the participating laboratories a better means of evaluating their performance and would eventually contribute to the improvement of cement testing.

A number of details were necessary for this preliminary study, which would not be required in a Standard Sample program. From the data reported in this study, the results of certain laboratories were consistently close to the means for all 103 laboratories, and the results of the others were about equally distributed around the means. Ten or 15 laboratories could be selected as referees from among the research, producer, commercial, State and Federal laboratories who have qualified by obtaining acceptable results. These laboratories could be used as a group to develop an average for each test for each sample. Since the chosen laboratories may obtain erratic results once in a while, provisions should be made to exclude values which appear to be out of line.

Cements for the standard samples would be carefully blended, then packaged and checked for uniformity, and mailed to those laboratories desiring to purchase them. Pairs of samples would be made available every two months, as in the program reported here, and on an annual basis.

The selected laboratories previously described would be requested to act as referee laboratories and submit their test results very promptly. The results of these referee laboratories would be averaged, and plots would be prepared using their means as the intersecting lines for scatter diagrams. As the results of those participating were received, they would be plotted on this diagram and, within 30 days or when all results were received, the scatter diagrams would be reproduced and distributed. The laboratories would immediately be able to see how their results compared with the accepted values and with those of all the other laboratories. The comparison would indicate to the participant where remedial action was necessary. A rating system such as described previously could also be adopted. Since the average values would appear on the scatter diagram, tolerances could be set up similar to those shown in Table IX and, with these two known, a laboratory could easily rate itself as was shown in Fig. 7. The rating would be a useful index for each laboratory to assess its conformance with a selected group of laboratories in the testing of cement.

Of course the ability to make these tests correctly is not a guarantee that a laboratory will make all other tests carefully or according to specification, but it certainly is a step in the right direction. If two laboratories were not in agreement in a given case, it would be possible to check their results on previous Standard Reference Samples to see if consistent differences were obtained. Purchasers of testing services may then assure themselves that the laboratories testing cement for them have acceptable ratings.

Laboratories whose results are not in agreement with the accepted values could obtain special attention and assistance from the Cement Reference Laboratory in their regular inspection service.

Scatter diagrams, such as those developed for this program, together with the participation of a large number of laboratories in a program, should offer the various working committees and sponsoring committees of ASTM Committee C-1 information as to whether or not the specification test methods are sufficiently clear and concise that operators can readily follow them.

SUMMARY

As a result of the tests by 103 laboratories on 12 samples of cement, it has been demonstrated that the use of paired samples, with graphic presentation as suggested by Youden, enables an appraisal of systematic laboratory errors and of the precision of the various tests. It was found that the standard deviation for many of the chemical and physical tests for cement was inflated by large differences between laboratories. The average standard deviations for the different tests on the 12 samples were usually about 2 to 3 times the estimated standard deviation attributable to the random errors of measurement associated with test methods.

There was a wide spread in the test results reported for both the physical and chemical tests. When the scores obtained by each laboratory were averaged over several samples or several tests, the number of laboratories obtaining high averages was larger than would be anticipated from a theoretical averaging of independent scores; and the same situation prevailed for the number of laboratories obtaining very low average scores.

The average values for the standard deviation for the different tests of this program were approximately the same as the corresponding average tolerance

values of 6 laboratories who regularly conduct interlaboratory tests. The percentage of participants meeting the tolerance requirements of these 6 laboratories ranged from 62 per cent to 95 per cent for the different tests.

Variables introduced into the program, such as (1) use of distilled water *versus* tap water for making cubes, (2) storage of the cement samples for 10 months in shipping bags in a storeroom or in sealed metal drums prior to distribution, and (3) special blending and packaging of chemical samples, did not have a statistically significant effect on average test values or the coefficient of variation of these values. The use of referee chemical methods did not result in better concordance than was obtained with the use of optional chemical methods.

A rating system for individual laboratories was developed which was based on the various test results reported by all the laboratories. Such a rating system, together with the scatter diagrams for each of the tests, enables a laboratory to evaluate its work and take remedial action where necessary.

Because of the apparent need for greater concordance in the testing of cement and because the preliminary reference sample program offered a satisfactory means for evaluating test procedures and laboratories, it has been proposed that a Cement Reference Sample program be established on a continuing basis.

APPENDIX I

LIST OF PARTICIPATING LABORATORIES

Alabama Highway Dept., Montgomery, Ala.
Alpha Portland Cement Co., Martins Creek, Pa., Birmingham, Ala.
Arizona Highway Dept., Phoenix, Ariz.
Ash Grove Lime & Portland Cement Co., Chanute, Kans.
Browser-Morner Testing Labs., Dayton, Ohio.
Calaveras Cement Co., San Andreas, Calif.
California Highway Dept., Sacramento, Calif.
California Portland Cement Co., Colton, Calif.
Colorado Highway Department, Denver, Colo.
Consolidated Cement Co., Fredonia, Kans.
E. L. Conwell & Co., Philadelphia, Pa.
Coplay Cement Mfg. Co., Coplay, Pa.
Corps of Engineers, Dallas, Tex., Jackson, Miss., Marietta, Ga.
Dewey Portland Cement Co., Dewey, Okla., Davenport, Iowa.
Dragon Cement Co., Northampton, Pa.
Diamond Portland Cement Co., Middle Branch, Ohio.
Froehling & Robertson, Inc., Richmond, Va.
General Portland Cement Co., Houston, Tex., Tampa, Fla.
Georgia Highway Dept., Atlanta, Ga.
Giant Portland Cement Co., Egypt, Pa.
Glens Falls Portland Cement Co., Glens Falls, N. Y.
Haller Testing Laboratories, Inc., Plainfield, N. J.
Halliburton Portland Cement Co., Corpus Christi, Tex.
Hercules Cement Corp., Nazareth, Pa.
Huron Portland Cement Co., Alpena, Mich.
Ideal Cement Co., Fort Collins, Colo., Baton Rouge, La.
Illinois Highway Dept., Springfield, Ill.
Indiana Highway Dept., Indianapolis, Ind.
Iowa Highway Dept., Ames, Iowa.
Kansas Highway Dept., Manhattan, Kans.
Keystone Portland Cement Co., Bath, Pa.
Kosmos Portland Cement Co., Kosmosdale, Ky.
Lehigh Portland Cement Co., Allentown, Pa., Birmingham, Ala.
Lone Star Cement Co., Hudson, N. Y., Dallas, Tex.
Louisiana Highway Dept., Baton Rouge, La.
Louisville Cement Co., Speed, Ind.
Maine Highway Dept., Orono, Me.
Marquette Cement Co., Chicago, Ill., Brandon, Miss.

Maryland State Roads Commission, Baltimore, Md.
Medusa Portland Cement Co., Wampum, Pa.
Manitowoc Portland Cement Co., Manitowoc, Wis.
Michigan State Highway Dept., Lansing, Mich.
Minnesota Highway Dept., St. Paul, Minn.
Missouri Highway Commission, Jefferson City, Mo.
Missouri Portland Cement Co., St. Louis, Mo., Independence, Mo.
The Monarch Cement Co., Humboldt, Kans.
Monolith Portland Cement Co., Monolith, Calif.
National Bureau of Standards, Allentown, Pa., Denver, Colo., San Francisco, Calif., Seattle, Wash., Washington, D. C.
National Cement Co., Ragland, Ala.
National Portland Cement Co., Bethlehem, Pa.
U. S. Naval Civil Engineering Laboratory, Port Hueneme, Calif.
Nazareth Portland Cement Co., Nazareth, Pa.
Nebraska Highway Dept., Lincoln, Nebr.
New Jersey Highway Dept., Trenton, N. J.
North Carolina Highway Dept., Raleigh, N. C.
The H. C. Nutting Co., Cincinnati, Ohio.
Ohio Highway Dept., Columbus, Ohio.
Olympia Portland Cement Co., Ltd., Bellingham, Wash.
Oregon Highway Dept., Salem, Ore.
Pacific Cement & Aggregates, Inc., Davenport, Calif.
Peerless Cement Co., Detroit, Mich., Port Huron, Mich.
Penn-Dixie Cement Corp., Nazareth, Pa., West Winfield, Pa.
Pennsylvania Highway Dept., Harrisburg, Pa.
Permanente Cement Co., Permanente, Calif., Lucerne Valley, Calif.
Pittsburgh Coke & Chemical Co., Neville Island, Pa.
Pittsburgh Testing Laboratory, Pittsburgh, Pa.
Portland Cement Assn. Chicago, Ill.
Bureau of Public Roads, Washington, D. C.
Raymond G. Osborne Laboratories, Inc., Los Angeles, Calif.
U. S. Bureau of Reclamation, Denver, Colo.
Standard Lime & Cement Co., Martinsburg, W. Va.
Shilstone Testing Laboratory, New Orleans, La.
Southern Cement Co., Birmingham, Ala.
Southwestern Portland Cement Co., Fairborn, Ohio, Victorville, Calif., El Paso, Tex.
Texas Highway Dept., Austin, Tex.
The Thompson & Lichtner Co., Inc., Brookline, Mass.
Toledo Testing Laboratory, Toledo, Ohio.
Twin City Testing & Engineering Lab., St. Paul, Minn.
Universal Atlas Cement Co., Gary, Ind., Northampton, Pa.
Volunteer Portland Cement Co., Knoxville, Tenn.
The Whitehall Cement Mfg. Co., Cementon, Pa.
West Virginia Highway Dept., Morganstown, W. Va.
Wisconsin Highway Dept., Madison, Wis.

APPENDIX II

AVERAGES OF RESULTS REPORTED BY ALL LABORATORIES

Cement Samples	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11	No. 12
Water consistency, per cent	24.6	25.5	27.0	25.9	24.9	27.1	28.7	24.5	24.4	25.6	26.0	25.9
Initial setting, Gillmore, hr	3.19	3.45	2.94	3.26	...	...	2.85	3.04	3.66	4.76	3.23	2.92
Final setting, Gillmore, hr	5.43	5.48	4.97	5.31	...	...	4.81	5.10	5.84	7.49	5.37	4.98
Expansion, per cent	0.14	0.13	0.06	0.08	0.03	0.04	0.02	0.18	0.03	0.03	0.09	0.06
Air entrainment, per cent	7.2	8.5	8.1	16.3	9.8	11.8	7.0	19.3	11.1	8.8	16.7	9.5
Water-air entrainment, per cent	70.3	69.5	70.0	60.5	68.5	67.0	70.4	57.5	67.1	67.8	58.8	68.5
Compressive strength, 3 days, psi	2327	2848	2764	2207	1889	2208	2675	1801	1868	2285	2292	2771
Compressive strength, 7 days, psi	3759	3675	4245	3617	2674	3645	4794	2481	2780	3389	3734	4302
Water for compressive strength, per cent	49.1	47.9	48.7	47.4	47.4	48.9	48.9	45.8	46.9	47.3	46.6	48.0
Fineness, Blaine, sq cm per g	3283	3442	3820	3107	3573	4564	4588	3147	3410	2949	3139	3875
SiO_2, per cent	21.08	21.73	21.74	21.70	21.24	23.22	20.76	22.25	21.65	23.48	21.65	21.68
Al_2O_3, per cent	6.83	4.95	5.87	5.88	5.47	8.13	5.60	5.12	5.76	3.67	5.90	5.90
Fe_2O_3, per cent	2.48	2.88	2.82	2.73	4.76	2.58	2.50	2.36	4.63	4.44	2.69	2.75
CaO, per cent	63.12	61.90	63.97	65.02	61.98	55.35	64.24	62.44	63.00	64.39	64.96	63.93
MgO, per cent	2.93	3.81	1.83	1.36	2.42	3.89	1.97	3.59	1.15	0.76	1.35	1.80
SO_3, per cent	1.80	2.79	1.82	1.78	1.92	2.43	2.84	2.18	1.87	1.80	1.78	1.82
Ignition loss, per cent	0.75	0.79	1.18	0.79	1.38	2.63	1.26	0.86	1.05	0.76	0.82	1.25
Residue, per cent	0.16	0.25	0.15	0.18	0.18	0.53	0.21	0.17	0.14	0.14	0.13	0.13
Na_2O, per cent	0.36	0.30	0.19	0.14	0.24	0.10	0.15	0.30	0.15	0.50	0.17	0.21
K_2O, per cent	0.70	1.00	0.74	0.75	0.59	0.25	0.76	1.08	0.82	0.21	0.77	0.76
Mn_2O_3, per cent	...	...	...	...	0.09	0.64	...	...	...	...	...	...
Sulfide sulfur, per cent	...	...	...	...	0.01	0.59	...	...	...	...	...	...
False setting, initial, mm	...	...	...	...	...	...	...	...	...	...	31	33
False setting, final, mm	...	...	...	...	...	...	...	...	...	...	20	19
Tensile strength, 3 days, psi	...	...	352	317	...	...	...	...	...	...	...	...
Tensile strength, 7 days, psi	...	...	424	395	...	...	...	...	...	...	...	...
Water for tensile strength, per cent	...	...	10.8	10.8	...	...	...	...	...	...	...	...
Vicat setting, hr	...	...	...	...	2.66	2.97	...	...	...	...	...	...
Fineness, Wagner, sq cm per g	...	...	...	...	...	...	2547	1739	...	...	...	...
Retained on No. 325 sieve, g	...	...	...	...	...	...	0.01	0.10	...	...	...	...
Sulfate resistance expansion, 14 days, per cent	...	...	...	...	...	...	...	...	0.005	0.002	...	...
Sulfate resistance expansion, 28 days, per cent	...	...	...	...	...	...	...	...	0.018	0.009	...	...
Heat of hydration, 7 days, cal per g	...	...	...	...	...	...	...	...	69	64	...	...
Heat of hydration, 28 days, cal per g	...	...	...	...	...	...	...	...	82	77	...	...

APPENDIX III

AVERAGE RESULTS, WITH COMPUTED STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR CHEMICAL ANALYSIS OF SAMPLES NO. 3 AND NO. 4.[a]

	Sample No. 3						Sample No. 4					
	Vial			Bulk			Vial			Bulk		
	Average, per cent	Standard Deviation	Coefficient of Variation, per cent	Average, per cent	Standard Deviation	Coefficient of Variation, per cent	Average, per cent	Standard Deviation	Coefficient of Variation, per cent	Average, per cent	Standard Deviation	Coefficient of Variation, per cent
Silicon dioxide (SiO_2)	21.73	0.183	0.84	21.75	0.182	0.84	21.71	0.215	0.99	21.69	0.174	0.80
Aluminum oxide (Al_2O_3)	5.88	0.241	4.10	5.86	0.315	5.38	5.88	0.165	2.81	5.90	0.310	5.25
Ferric oxide (Fe_2O_3)	2.84	0.266	9.37	2.80	0.171	6.11	2.74	0.257	9.38	2.73	0.138	5.05
Calcium oxide (CaO)	63.91	0.447	0.70	64.03	0.280	0.44	65.02	0.348	0.54	64.98	0.353	0.54
Magnesium oxide (MgO)	1.84	0.138	7.50	1.81	0.156	8.62	1.36	0.162	11.91	1.34	0.106	7.91
Sulfur trioxide (SO_3)	1.82	0.071	3.90	1.81	0.104	5.75	1.76	0.093	5.28	1.79	0.083	4.64
Loss on ignition	1.16	0.133	11.47	1.20	0.141	11.75	0.80	0.163	20.38	0.80	0.164	20.50
Insoluble residue	0.13	0.049	37.69	0.17	0.110	64.70	0.18	0.165	91.67	0.14	0.108	77.14
Sodium oxide (Na_2O)	0.18	0.024	13.33	0.19	0.031	16.32	0.14	0.032	22.86	0.15	0.022	14.67
Potassium oxide (K_2O)	0.75	0.038	5.28	0.73	0.053	7.26	0.74	0.045	6.08	0.76	0.023	3.03

[a] Half of the laboratories made tests on specially blended samples of cement packaged in sealed vials, whereas the other laboratories used portions of cement from the bulk samples for chemical analysis.

AVERAGE RESULTS, WITH COMPUTED STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR COMPRESSIVE STRENGTH TESTS OF SAMPLES NO. 7 AND NO. 8.[a]

	Sample No. 7						Sample No. 8					
	Distilled			Tap			Distilled			Tap		
	Average, psi	Standard Deviation	Coefficient of Variation, per cent	Average, psi	Standard Deviation	Coefficient of Variation, per cent	Average, psi	Standard Deviation,	Coefficient of Variation, per cent	Average, psi	Standard Deviation	Coefficient of Variation, per cent
1 day	2644	190	7.2	2708	204	7.5	...	...	...	...	...	...
3 days	4796	262	5.5	4777	259	5.4	1799	139	7.7	1788	139	7.8
7 days	...	...	...	...	...	...	2442	161	6.6	2429	185	7.6

[a] Half the laboratories made the test by regular ASTM procedure, while the other laboratories used distilled water for making cubes, after which the cubes were stored in a moist plastic bag in the moist cabinet until time for breaking.

[Continued on page 1154.]

AVERAGE RESULTS, WITH COMPUTED STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR ALL THE PHYSICAL AND CHEMICAL TESTS ON SAMPLES NO. 11 AND NO. 12.[a]

	Sample No. 11						Sample No. 12					
	Skid			Drum			Skid			Drum		
	Average	Standard Deviation	Coefficient of variation	Average	Standard Deviation	Coefficient of Variation	Average	Standard Deviation	Coefficient of Variation	Average	Standard Deviation	Coefficient of Variation
Water consistency, per cent	25.4	0.504	1.99	25.8	0.543	2.10	25.6	1.006	3.93	26.0	0.550	2.12
Initial setting, Gillmore, hr	3.21	0.251	7.82	3.24	0.395	12.19	2.91	0.418	14.35	2.93	0.485	16.55
Final setting, Gillmore, hr	5.47	0.716	13.09	5.27	0.654	12.41	4.88	0.584	11.98	5.07	0.681	13.43
Expansion, per cent	0.071	0.018	24.7	0.076	0.019	25.0	0.051	0.013	24.6	0.059	0.018	35.3
Air entrainment, per cent	16.6	1.21	7.26	16.8	1.80	10.72	9.46	1.17	12.33	9.59	1.37	14.31
Water-air entrainment, per cent	58.3	2.39	4.11	59.3	3.13	5.27	68.6	2.11	3.07	68.4	2.08	3.04
Compressive strength 3 days, psi	2346	184	7.84	2229	222	9.96	2788	292	10.47	2767	207	7.48
Compressive strength 7 days, psi	3787	321	8.48	3674	251	6.83	4352	292	6.71	4256	340	7.99
Water for compressive strength, per cent	46.2	0.993	2.15	46.9	1.12	2.40	47.9	1.27	2.64	48.0	0.912	1.90
Fineness, Blaine, sq cm per g	3153	66.0	2.09	3125	180	5.76	3892	121	3.11	3859	94	2.44
False setting, initial mm	34.15	3.21	9.40	27.71	4.81	17.36	34.09	5.87	17.22	33.60	5.56	16.55
False setting final mm	22.21	4.93	22.20	17.65	4.97	28.16	12.88	5.84	45.34	25.28	6.42	25.40
SiO_2, per cent	21.66	0.305	1.41	21.64	0.378	1.75	21.65	0.310	1.43	21.70	0.176	0.81
Al_2O_3, per cent	5.88	0.18	3.06	5.92	0.251	4.24	5.91	0.228	3.86	5.89	0.248	4.21
Fe_2O_3, per cent	2.71	0.071	2.60	2.68	0.119	4.44	2.74	0.114	4.16	2.76	0.063	2.28
CaO, per cent	64.95	0.419	0.65	65.00	0.495	0.76	63.96	0.367	0.57	63.91	0.460	0.72
MgO, per cent	1.34	0.127	9.48	1.39	0.220	15.83	1.79	0.231	12.91	1.82	0.185	10.17
SO_3, per cent	1.78	0.059	3.33	1.78	0.059	3.33	1.83	0.833	4.55	1.81	0.081	4.47
Ignition loss, per cent	0.86	0.104	12.09	0.77	0.097	12.58	1.28	0.140	10.94	1.23	0.129	10.49
Residue, per cent	0.13	0.060	45.76	0.13	0.063	48.54	0.13	0.056	43.00	0.13	0.472	36.31
Na_2O, per cent	0.17	0.020	11.88	0.17	0.042	24.53	0.21	0.043	20.67	0.21	0.022	10.52
K_2O, per cent	0.77	0.031	3.99	0.76	0.056	7.33	0.76	0.050	6.61	0.76	0.037	4.80

[a] Half of each of these samples packaged for shipment were stored 10 months in sealed metal drums and the other half were stored 10 months on a platform skid in a storeroom.