THE PRECISION OF SELECTED AGGREGATE TEST METHODS

Paul E. Benson and W. H. Ames, Transportation Laboratory,
California Department of Transportation

The interlaboratory correlation program pilot study is briefly discussed. Precision statements derived from the study are presented for the following test methods: sieve analysis, percent crushed particles, L.A. rattler, sand equivalent, cleanness value, durability index, and R-value. Relative amounts of general error types such as between operator and between laboratory are given for each test method, and possible causes are discussed. Laboratory performance is shown through the use of scatter diagrams and ranking summaries. Recommendations for improving test precision are given.

•OVER the years a number of valuable test methods have been developed for judging the quality of aggregate used in portland cement concrete, asphalt concrete, and base and subbase construction. When applied properly, these tests have been used consistently to accept material of adequate quality and to reject material of inferior quality. Until recently, however, only a minimum effort has been made to measure and improve the precision of these tests. Active calibration and certification programs have sought to identify testing errors so that they might be reduced. However, these programs have been handicapped by lack of knowledge about the magnitude and source of these errors. An integrated method for continually monitoring test precision and evaluating laboratory and operator performance is needed.

This report summarizes the results of a year-long pilot study that measured the precision of a number of aggregate test methods, quantified the sources of testing error, and evaluated laboratory performance. The test methods studied were coarse and fine sieve analysis, R-value, L.A. rattler abrasion, fine durability, coarse durability, cleanness value, and percent crushed particles. The precision of the sand equivalent test, determined under a separate study (4), is also included in this report.

The results contained herein were analyzed by a series of computer programs developed especially for this study. These programs are fully explained in another report (1).

To clarify some of the conclusions reached in this report, a discussion of the concepts of precision and testing error is necessary. California has adopted a method of reporting test precision recommended in ASTM Designation C 670-71T. This is based on a statistical parameter called the difference 2-sigma limit (D2S). ASTM uses the D2S limit to form 2 different types of precision statements:

1. Single-operator precision—a measure of the greatest difference between 2 results that would be considered acceptable when properly conducted determinations are made on uniformly prepared portions of material by a competent operator using 1 set of equipment.
2. Multilaboratory precision—a measure of the greatest difference between 2 results that would be considered acceptable when properly conducted determinations are made by 2 different operators in different laboratories on uniformly prepared portions of material.

Publication of this paper sponsored by Committee on Quality Assurance and Acceptance Procedures.

Single-operator and multilaboratory precision statements are given in this report for each test method studied. The D2S limit is referred to as the "acceptable range of 2 results" in these statements. For many of the tests, precision was found to vary significantly according to the range of material tested. The precision statement is given in a tabular form for these test methods. The overall range of material studied for each test method is also given. Precision statements are accurate for this range only and should not be extrapolated.

Testing error was divided into 2 general categories for the purposes of this study. The first, systematic error, is composed of errors whose sources are identifiable. For this experiment the identifiable sources of error were between laboratories, between operators in the same laboratory, and scale-type error (3). A large between-laboratory error might indicate significant variations from laboratory to laboratory in technique, environment, or equipment. A large between-operator error could indicate inadequate training and certification programs at the local level. Scale-type errors are caused by inconsistencies between expected and observed test results from one range of results to another. Significant scale-type errors usually occur in test methods that use different equipment or techniques for each range of material tested. For instance, a set of poorly calibrated standard weights would yield a large scale-type error when weighing objects of varying sizes. Systematic errors can often be minimized because their causes are usually known.

The second type of testing error, residual error, represents the total of all errors not accounted for by the systematic components of operator, laboratory, and scale-type effects. Minimizing this type of error can be more difficult. If additional experimentation does not reveal more systematic components of the residual error, the precision of the final result can only be improved by averaging a predetermined number of repeated tests for each test result or by tightening method and equipment tolerances. Before this is done, however, the magnitude of the sample-preparation error (a measure of uniformity of the sample-preparation procedure) should be checked. If this error is a large part of the overall residual error, then the actual test precision will be better than indicated and may not need improvement.

Single-operator precision was calculated from the residual error, and as such included the random errors inherent in both the test method and the sample-preparation procedure. Multilaboratory precision was derived from a combination of systematic and residual errors and therefore included effects of laboratory environments, equipment, and operator technique in addition to the residual error.

DESCRIPTION OF WORK

The California Department of Transportation's 11 district materials laboratories and its headquarters laboratory were the participants in this pilot study. Sample preparation and data analysis were handled by Transportation Laboratory personnel in Sacramento. The testing program was spread out over almost 2 years, and the analysis phase, speeded by the use of the computer, was completed in several months.

The samples were prepared and distributed in sets of 2. Samples in each set were of the same aggregate type (i.e., AC, PCC, AB, or AS) but were obtained from 2 different sources. The test methods performed on each set of samples are given in Table 1. A total of 10 individual samples were studied. The total amount of testing to be done was determined by theoretical design considerations tempered by practical constraints.

Each sample was split into 64 subsamples. Each of these subsamples contained enough material to perform 1 series of tests, and 48 of these were randomly assigned to the different laboratories. The remaining 16 subsamples were kept as a contingency. Thus, each of the 12 participating laboratories received 4 subsamples from each sample (Figure 1).

At the beginning of each 3-month interval the laboratories received their 2 sets of 4 subsamples each. They then chose 2 operators and set aside 1 set of equipment. On the day or days that the tests were to be made, each operator was given 2 subsamples

Table 1. Summary of testing.

Samples	Aggregate Type	Dates Tested	Tests Studied	Calif. Test Method No.
1 and 2	PCC (fine)	3/72-5/72	Fine sieve analysis	202G
			Fine durability index	229E
3 and 4	PCC (coarse)	3/72-5/72	Coarse sieve analysis	202G
			L.A. rattler	211D
			Cleanness value	227E
5 and 6	AB	6/72-10/72	Sieve analysis	202G
			Durability index	229E
			Percent crushed particles	205E
			R-value	301F
7 and 8	AC	7/73-11/73	Sieve analysis	202G
			L.A. rattler	229E
			Percent crushed particles	205E
9 and 0	AS	4/73-6/73	Sieve analysis	202G
			R-value	301F

Figure 1. Sample distribution.

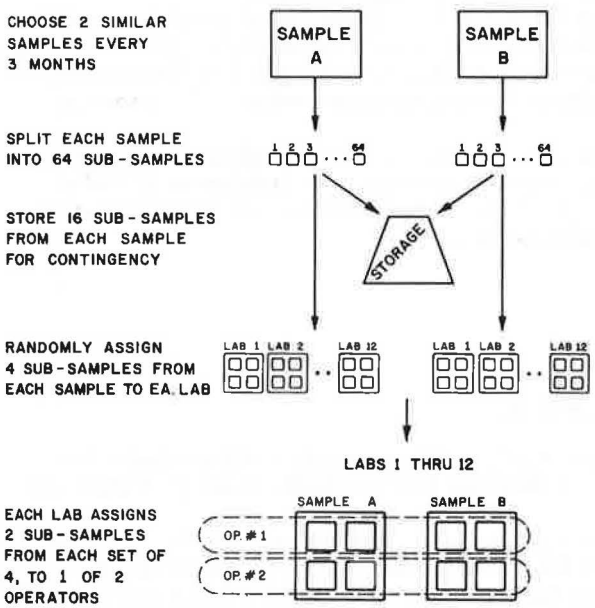
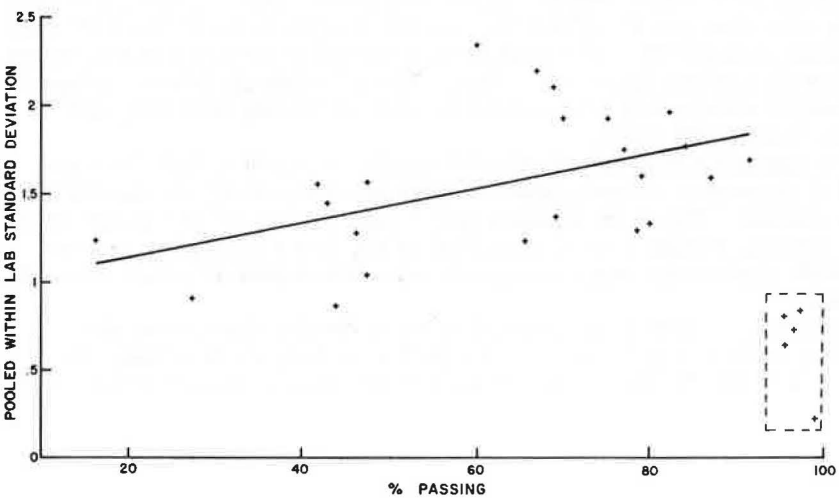


Figure 2. Test precision versus material range, coarse sieve analysis.



from each sample. The operators ran the indicated tests following their usual procedure. The operators used the same set of equipment for the 4 subsamples tested.

ANALYSIS

Only brief mention of the analytical techniques employed in this study will be made here. A more complete discussion can be found elsewhere (1).

The relationship between test precision and material range was investigated first for each of the test methods studied. This was done by linearly regressing the pooled within-lab standard deviation against the overall sample average. If a significant relationship was indicated, data transformation was required before further analysis could be made.

Precision statements were determined by using a 3-factorial analysis of variance and isolating the components of variance according to expected mean-square equations. These same components were used to estimate the relative distribution of the general error types: between-operator, between-laboratory, and residual.

Cell variances were not homogeneous for the test methods whose precision varied with material range. Because overall homogeneity of cell variance is a prime assumption on which the theory of analysis of variance relies, logarithmic transformations were used extensively. Results were then retransformed according to the common rules for propagation of error (2).

Scale-type errors were derived from Mandel's linear model analysis (2, 3). Variations in the distribution of errors from different sources as a function of material range were also studied using this method. Laboratory performance was monitored by using scatter diagrams and ranking summaries.

FINDINGS AND CONCLUSIONS

Sieve Analysis (Test Method No. Calif. 202-G)

The sieve analysis test method is divided into 2 parts—a coarse analysis and a fine analysis. Because these are, in effect, 2 different test methods, their precision was studied separately.

The coarse analysis procedure is used for material retained on the No. 4 and coarser sieves. Test precision for these sieves was found to be roughly dependent on the total weight of material passing them. Except for the range of 95 to 99 percent passing, the greater the weight of material passing a coarse sieve, the less repeatable were its results. Apparently, shaking time became more critical and errors from sieve defects were magnified as a greater weight of material passed through a given sieve. The dependent nature of one sieve result on another makes this impossible to prove, however. The assumption was made for this study that the percent passing a sieve was a reasonably consistent representation of the actual weight of material passing the sieve, since sample sizes were fairly uniform from test to test. The relationship between percent passing and repeatability should only be considered a rule of thumb, however, and should not be applied in extreme cases.

Figure 2 shows the pooled within-lab standard deviation (a result of both between-operator and residual sources of error) plotted against percent passing for all coarse sieve-sample combinations. The least squares linear plot shown, which was not based on 95 to 99 percent passing results (shown in dashed area), has a coefficient of correlation of 0.49. Table 2 gives the single-operator and multilaboratory precision of the coarse sieve analysis.

A fine analysis procedure is used for material passing the No. 4 and finer sieves. This method combines hydraulic and mechanical agitation techniques to gradate the sample and wash out clay and silt particles. Table 3 gives its precision over the range studied.

The fine sieve results are weighted according to the amount of material passing the No. 4 sieve to yield combined or overall results for sieves No. 8 through No. 200. Figure 3 shows the pooled within-lab standard deviation plotted against the percent passing for these results. The coefficient of correlation for this linear regression is 0.89. Table 4 summarizes the precision of the combined sieve analysis.

The most dominant source of error for both the coarse and fine sieve analyses was residual error. It is presumed that the largest part of this error was caused by the inability to accurately split samples into identical subsamples.

Percent Crushed Particles Retained on No. 4 Screen (Test Method No. Calif. 205-E)

The relative amount of crushed material contained in a sample of aggregate is evaluated by inspection. The 4 samples tested by this method ranged from approximately 55 percent to 95 percent crushed particles. The test exhibited very large systematic errors, particularly between laboratories. The error distribution was as follows: between laboratory, 65 percent; between operator, 20 percent; and residual error, 15 percent. The precision of the crushed particle test was shown to be very poor, especially for materials with low crushed-particle counts (Table 5). Discrepant results roughly correlated with geographical location, with laboratories in the southern part of California getting significantly lower results than the rest of the state.

The large errors measured for this test method are most likely caused by the highly subjective nature of the test. If this test is to be used as a contract control test, the source of these errors must be identified and minimized.

L.A. Rattler (Test Method No. Calif. 211-D, 500 Rev.)

The L.A. Rattler Test is used to measure the resistance of coarse aggregate to degradation caused by impact. The range of results studied for this test method was 13 to 18 percent loss. The precision measured was constant over this range, as shown in the following precision statement:

Type	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator	1.10	1.05	3.0 percent loss
Multilaboratory	3.53	1.88	5.3 percent loss

An analysis of the components of variance revealed that between-laboratory error constituted 70 percent of the overall error. Residual error made up the remaining 30 percent, while between-operator error was negligible. Since each laboratory has only 1 Los Angeles Abrasion Testing Machine, it becomes obvious that equipment, not operator technique, is the most critical factor affecting the precision of the test.

Sand Equivalent (Test Method No. Calif. 217-I)

The precision of the sand equivalent test method was determined and reported under a separate study (10) and is included here for completeness. The single-operator precision was as follows:

Table 2. Precision statement tabulation, coarse sieve analysis (¾-in. through No. 4).

Percent Passing	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
20	1.09	1.04	3.0
40	1.49	1.22	3.5
60	1.96	1.40	4.0
80	2.48	1.58	4.5
1 to 5 and 95 to 99	0.56	0.75	2.1
Multilaboratory precision			
20	1.58	1.26	3.6
40	2.16	1.47	4.2
60	2.83	1.68	4.8
80	3.59	1.89	5.4
1 to 5 and 95 to 99	1.17	1.08	3.1

Table 3. Precision statement tabulation, fine sieve analysis (No. 8 through No. 200).

Percent Passing	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
20	1.11	1.06	3.0
40	1.95	1.40	4.0
60	3.02	1.74	4.9
80	4.32	2.08	5.9
Multilaboratory precision			
20	1.68	1.30	3.7
40	2.95	1.72	4.9
60	4.57	2.14	6.0
80	6.54	2.56	7.2

Figure 3. Test precisions versus material range, combined sieve analysis.

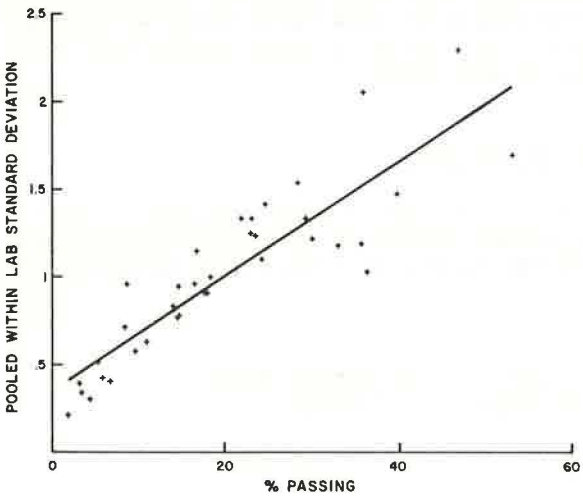


Table 4. Precision statement tabulation, combined sieve analysis (No. 8 through No. 200).

Percent Passing	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
10	0.41	0.64	1.8
20	0.91	0.95	2.7
30	1.60	1.26	3.6
40	2.48	1.57	4.5
50	3.55	1.89	5.3
Multilaboratory precision			
10	0.56	0.75	2.1
20	0.22	1.11	3.1
30	2.15	1.47	4.1
40	3.34	1.83	5.2
50	4.78	2.19	6.2

Table 5. Precision system tabulation, percent crushed particles (retained on No. 4).

Percent Crushed	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
60	30.74	5.54	16
70	20.02	4.47	13
80	11.59	3.40	10
90	5.45	2.33	7
Multilaboratory precision			
60	218.90	14.80	42
70	142.56	11.94	34
80	82.52	9.08	26
90	38.80	6.23	18

Sand Equivalent Range	Variance	Standard Deviation	Acceptable Range of 2 Results
Below 45	1.87	1.37	3.9
45 to 65	8.72	2.95	8.4
Above 65	4.27	2.07	5.9

The multilaboratory precision was as follows:

Sand Equivalent Range	Variance	Standard Deviation	Acceptable Range of 2 Results
Below 45	2.90	1.70	4.8
45 to 65	14.05	3.75	10.6
Above 65	7.03	2.65	7.5

Cleanness Value (Test Method No. Calif. 227-E)

The cleanness test indicates the amount, fineness, and character of clay-like materials and coatings present in coarse aggregate. Precision of the test was based on 2 samples in the 90 to 95 percent cleanness value range. The conclusions drawn from the limited data are preliminary and will be augmented in the future by a continuous correlation program that has already been implemented.

Between-operator error was found to be insignificant, whereas between-laboratory error constituted over 40 percent of the total error. This tends to indicate that there are either equipment calibration deficiencies or lack of uniform application of testing procedures from laboratory to laboratory. The actual errors are reasonably small, however, as illustrated by the precision statement:

Type	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator	0.69	0.83	2.3 CV units
Multilaboratory	1.21	1.10	3.1 CV units

Durability Index (Test Method No. Calif. 229-E)

The durability index is a measure of an aggregate's resistance to producing detrimental clay-like fines when subjected to certain chemical and mechanical forms of degradation. Both fine and coarse durability methods are used. The precision of the 2 methods is given in Tables 6 and 7 respectively. Test precision improves with increased durability.

Since coarse durability was measured for only 2 samples, the precision measurements in Table 7 should be considered preliminary. However, fine durability results were recorded for 4 samples, permitting fairly reliable measurement of the systematic errors. The breakdown of the overall fine durability error was as follows: between laboratory, 50 percent; between operator, 30 percent; and residual error, 20 percent. For high-range material, however, between-laboratory error diminished to 20 percent, whereas for low-range material it increased to 60 percent. This indicates that the test is more sensitive at low durabilities than high durabilities to some source of error occurring between the laboratories. This error could be caused by differences

Table 6. Precision statement tabulation, fine durability.

Fine Durability	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
50	5.74	2.40	6.8
60	4.33	2.08	5.9
70	3.11	1.76	5.0
Multilaboratory precision			
50	26.07	5.11	14.4
60	19.65	4.43	12.5
70	14.14	3.76	10.6

Table 7. Precision statement tabulation, coarse durability.

Coarse Durability	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
60	12.85	3.58	10.1
70	6.53	2.56	7.2
80	2.33	1.53	4.3
Multilaboratory precision			
60	18.88	4.35	12.3
70	9.59	3.10	8.8
80	3.42	1.85	5.2

Table 8. Precision statement tabulation, R-value.

R-Value	Variance	Standard Deviation	Acceptable Range of 2 Results
Single-operator precision			
30	38.54	6.21	18
40	27.87	5.28	15
50	18.92	4.35	12
60	11.69	3.42	10
70	6.20	2.49	7
80	2.43	1.56	4
Multilaboratory precision			
30	76.40	8.74	25
40	55.24	7.43	21
50	37.49	6.12	17
60	23.18	4.81	14
70	12.29	3.51	10
80	4.83	2.20	6

in calcium chloride solutions, tap water, temperature control, or agitators. Further study to identify which of these factors significantly affects the precision and then eliminating that error should substantially improve the precision of the test.

The 2 sets of samples on which fine durability measurements were made were sent out 3 months apart. For the most part, the same operators ran the tests using the same equipment. However, a significant within-laboratory scale-type error was measured. It appears that the most probable source of this error was a change either in laboratory temperature or calcium chloride solution concentration during the 2-month period.

R-Value (Test Method No. Calif. 301-F)

The 4 samples tested ranged in R-value from 30 to 85. As with many of the other tests, precision was found to vary according to the range of material tested. In the range tested, low R-value material yielded less precise test results than high R-value material.

Table 8 summarizes the single-operator and multilaboratory precision for the R-value test. The overall distribution of errors was as follows: between laboratory, 30 percent; between operator, 20 percent; and residual error, 50 percent. Between-laboratory error was greater than the 30 percent listed for low-range material. Also, significant scale-type errors of both the within- and between-laboratory varieties were observed. The scale-type errors were possibly caused by stabilometer readings, since these instruments, if not properly calibrated, can give high results in one range and low results in another. Intricate sample fabrication procedures probably contributed to a large portion of the residual error measured.

ACKNOWLEDGMENTS

Without the hard work of the numerous laboratory technicians throughout the state, this report would not have been possible. The authors wish to thank the Pavement Branch of the Transportation Laboratory, particularly Thomas Whitney, for performing the cumbersome and tedious job of splitting the bulky samples into manageable and homogeneous subsamples. Also, special thanks go to the laboratory personnel for diligently carrying out the extensive testing program in each of the 11 transportation districts.

This project was conducted in cooperation with the Federal Highway Administration. The contents of this report reflect the views of the Transportation Laboratory, which is responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

REFERENCES

1. Paul E. Benson. Development of a Correlation Program. California Department of Transportation, Report CA-DOT-TL-1153-4-75-06, 1974.
2. John Mandel. The Measuring Process. Technometrics, Aug. 1959, pp. 251-267.
3. John Mandel and T. W. Lashof. The Interlaboratory Evaluation of Testing Methods. ASTM Bulletin, No. 239, July 1959, pp. 53-61.
4. R. R. Svetich and P. E. Benson. Precision of the Sand Equivalent Test. California Department of Transportation, Report TL-1153-1-74-06, 1974.