A TEST OF THE PRECISION OF THIN-SECTION ANALYSIS BY POINT COUNTER

F. Chayes, *Geophysical Laboratory, Washington, D. C.*

AND

H. W. Fairbairn, *Massachusetts Institue of Technology, Cambridge, Mass.*

Abstract

Five thin sections of granite from Westerly, Rhode Island, were analyzed by each of five relatively inexperienced operators using the point-counter method of determining the mode. Each operator ran his slides in a different sequence. In a preliminary run the observed analytical error for the constituents exceeded the expected error; in a second run by the same operators, now more experienced, the observed analytical error was approximately the same as the expected error. Evaluation of variances showed order of analysis to be insignificant throughout, operator and slide differences to be slight. The means for all constituents in both runs were remarkably close to the means previously obtained by Chayes in an independent run on the same sections. The test definitely establishes the level of precision for the point-counter method of analysis on a broader base than was possible from Chayes' unchecked analyses. The convenience and rapidity of the method are already well known; its precision has now been thoroughly tested.

The point counter described in this journal some time ago (Chayes, 1949) is now rather widely used and its popularity seems to be increasing steadily. The only published test of the precision of the method, that which accompanied the original description of the instrument, contained no information at all about interoperator differences, for the reason that at that time there were no other operators. In general, results obtained by different operators may disagree because of differences in identification or tabulation conventions, or because of differences in technique or competence. In the first case, and also where both types of difference are present, the results will be subject to a bias which varies from operator to operator. Where a difference in competence only is involved, the results are to be thought of as drawn from parent populations having the same means but different variances. We rarely have enough information to judge which type of difference is responsible for divergences between published results, and it is usually safe to assume the presence of both, as well as of some difference between samples.

It is pleasant to be able to report that the earlier precision test was apparently sound, that for major constituents the error attributable solely to the analytical process evidently is binomial,[1] as alleged, and,

---

[1] The "expected" or theoretical analytical error, as a number of points, is given by $\sqrt{np(1-p)}$, where p is the percentage of constituent and n the (total) count length. On a percentage basis this becomes $\sqrt{p(1-\mathrm{p})/n}$. The coefficient of variation, which is sometimes used as an index of relative error, is then $C=\sqrt{((1-p)/np}$.

most important, that interoperator differences can easily be held within reasonable limits. Results warranting these assertions have been obtained at Massachusetts Institute of Technology by a group consisting of C. K. Bell, W. H. Dennen, H. W. Fairbairn, M. L. Jensen, and F. B. Whiting.

A preliminary test was run by these analysts in May, 1950, using five thin sections cut from a single strip of Westerly granite. On a point counter similar to that in use at the Geophysical Laboratory each man analyzed each section once. The schedule followed by the analysts is shown in Table 1, in which the slides are identified by letters.

TABLE 1. SCHEDULE OF TEST

Order	1	2	3	4	5
Analyst I	A	B	C	D	E
II	E	A	B	C	D
III	D	E	A	B	C
IV	C	D	E	A	B
V	B	C	D	E	A

The results of this test were promising rather than satisfying. Agreement between the mean values obtained at MIT and at the Geophysical Laboratory was astonishingly good (see Table 2), and from this it was clear that there were no major identification biases. (The Geophysical Laboratory analyses were made some time before the test was started but the results were not known to any of the participants.)

Detailed examination of the results of the first run revealed some minor differences of opinion as to identification and some confusion about tabulation conventions. Finally, and most surprising in view of the excellent check of mean values, the observed analytical errors were significantly larger than predicted values for two of the three major constituents. (Compare rows 1 and 2 of Table 3.) These matters were discussed by the group and it was pointed out to them that the excess analytical error might be attributable to differences in both competence and learning rates, particularly the latter, if the practice period before the test had been inadequate.

Six months later the test was repeated with results shown in Tables 2 and 4*B* and in rows 3 and 4 of Table 3. In this run significant excess of observed over expected analytical error is not in evidence in the major constituents. It is still present in muscovite and for biotite it is larger in the second than in the first run. When results of the first test were compared with claims made earlier for the method (see Chayes, 1949, p. 7) it seemed entirely possible that the theory invoked seriously over-

TABLE 2. MEAN VALUES OBTAINED AT MIT AND GEOPHYSICAL LABORATORY*
(Volume Per Cent)

Mineral	Analyst Averages, MIT					Grand Means	
	I	II	III	IV	V	MIT	Geophysical Laboratory
Quartz	26.1 26.8	26.2 28.3	26.3 26.3	27.8 27.2	27.3 27.4	26.8 27.2	27.1
Microcline	36.1 35.0	34.9 35.0	35.7 35.7	35.0 35.9	35.7 34.5	35.5 35.2	35.0
Plagioclase	32.4 32.6	33.7 31.6	32.7 32.0	29.7 31.8	30.7 32.7	31.8 32.1	32.2
Biotite	3.1 3.3	2.8 2.8	3.0 4.1	2.3 2.7	3.1 2.9	2.9 3.2	3.4
Muscovite	0.9 1.2	1.3 1.3	1.1 1.1	3.1 1.0	2.0 1.3	1.7 1.2	1.2
Opaques	0.8 0.7	0.7 0.7	0.8 0.9	1.0 0.9	0.9 0.7	0.9 0.8	0.7
Nonopaques	0.5 0.4	0.4 0.3	0.3 0.2	1.1 0.6	0.3 0.5	0.5 0.4	0.5

* Upper entry at left is for first MIT test, lower entry on right for second.

estimated precision. For major and accessory constituents this possibility is clearly eliminated by the second test. It is probable, therefore, that with a little further practice this group of analysts would soon be obtaining results in accord with prediction for all constituents. The only remaining difficulty seems to be inability to stabilize identification conventions for the micas, and this difficulty may be anticipated whenever

TABLE 3. OBSERVED AND EXPECTED ANALYTICAL ERROR IN THE MIT TESTS*
(Per Cent of Whole)

Mineral	Quartz	Microcline	Plagioclase	Biotite	Muscovite	Opaque	Nonopaque
First Test							
Observed	1.3	1.4	1.1	0.4	0.4	0.2	0.2
Expected	0.88	0.96	0.93	0.33	0.26	0.18	0.14
Second Test							
Observed	1.0	1.1	0.9	0.6	0.4	0.1	0.1
Expected	0.89	0.96	0.93	0.35	0.22	0.17	0.12

* See footnote 1.

biotite is partly replaced by chlorite and much of the muscovite occurs in minute flakes replacing plagioclase.

TABLE 4*A*. INDIVIDUAL RESULTS, FIRST MIT TEST
(Volume Per Cent)

Analyst	Slide	Quartz	Microcline	Plagioclase	Biotite	Muscovite	Opaques	Nonopaques
I	A	25.7	35.8	33.3	2.6	1.2	1.0	0.4
	B	25.5	38.0	32.3	2.4	0.6	0.7	0.5
	C	26.8	35.4	31.8	3.6	0.8	1.0	0.6
	D	25.4	38.0	31.4	3.0	0.6	1.0	0.7
	E	27.0	33.4	33.2	3.9	1.3	0.5	0.6
II	A	24.8	35.0	34.4	2.9	1.6	0.7	0.5
	B	26.2	35.4	32.9	3.0	1.3	0.8	0.4
	C	26.2	35.7	34.0	2.4	0.8	0.7	0.2
	D	25.8	35.9	33.7	2.3	1.3	0.6	0.4
	E	28.1	32.5	33.3	3.2	1.7	0.7	0.5
III	A	24.6	36.5	33.3	3.4	1.3	0.6	0.3
	B	25.7	35.2	33.7	3.2	1.2	0.8	0.2
	C	27.2	36.6	32.0	2.7	0.4	0.9	0.2
	D	25.0	37.9	31.8	2.8	1.2	0.8	0.5
	E	29.2	32.3	32.7	3.0	1.2	1.0	0.5
IV	A	27.5	35.2	28.7	2.4	4.1	1.0	1.1
	B	28.4	34.2	30.3	2.6	2.9	0.8	0.8
	C	28.2	35.1	29.8	1.7	2.4	1.6	1.3
	D	26.9	34.6	30.8	2.5	3.0	1.0	1.0
	E	28.0	35.9	28.7	2.5	3.1	0.8	1.0
V	A	29.0	33.9	29.8	3.7	2.5	1.0	0.2
	B	26.5	38.2	30.2	2.7	1.3	0.6	0.4
	C	27.5	36.5	29.8	2.6	2.4	1.0	0.4
	D	28.2	35.2	30.2	3.0	1.9	1.1	0.4
	E	25.4	34.9	33.3	3.6	1.8	0.9	0.1

A brief note on analytical error may be helpful to some readers. It is important to remember that this error has nothing to do with mistakes or differences of opinion about identification, and expresses uncertainty inherent in the method under optimum circumstances. On the assumption that all five operators are competent and that each is using a stable set of identification conventions, the schedule of calculations is such as to throw differences ordinarily regarded as "mistakes" into the operator and slide variances (see Table 5). The remaining (analytical error) variance then expresses the failure of all results to agree exactly *after due allowance has been made for differences between both slides and operators*. It is part of the method, and if uncertainty arising from it is too large for comfort this uncertainty can be reduced only by extending the count length (e.g., reducing the vertical or horizontal distances between clicks) or by resorting to replicate analyses. For the individual worker the first of these alternatives is always preferable because the number of random replications which can be made on the same slide by the same operator-instrument combination is very small, and nothing is gained by retracing one's steps exactly.

If the error of the method is actually binomial, as alleged, entries in

TABLE 4*B*. INDIVIDUAL RESULTS, SECOND MIT TEST
(Volume Per Cent)

Analyst	Slide	Quartz	Micro-cline	Plagio-clase	Biotite	Musco-vite	Opaques	Non-opaques
I	A	24.7	35.6	33.3	3.3	2.0	0.6	0.6
	B	26.8	35.7	32.6	3.5	0.4	0.6	0.4
	C	28.0	34.2	32.1	3.4	1.1	0.7	0.4
	D	27.8	35.0	31.5	3.3	1.0	0.9	0.5
	E	26.6	34.5	33.6	3.0	1.4	0.6	0.3
II	A	27.3	35.5	32.1	2.5	1.5	0.8	0.3
	B	27.3	35.4	31.7	3.4	1.4	0.6	0.1
	C	28.0	35.3	31.1	2.8	1.4	0.8	0.5
	D	30.1	33.8	31.5	2.6	0.9	0.7	0.2
	E	28.7	35.2	31.4	2.6	1.3	0.6	0.2
III	A	25.8	36.0	33.5	2.9	0.8	0.7	0.3
	B	25.5	33.9	33.7	4.9	0.8	1.0	0.1
	C	26.1	37.8	30.7	3.4	1.1	0.7	0.3
	D	26.2	36.0	29.5	5.7	1.3	1.1	0.1
	E	27.8	34.7	32.4	3.6	0.7	0.8	0.2
IV	A	26.4	36.2	32.7	2.1	1.1	0.8	0.6
	B	26.6	36.3	31.9	3.2	0.8	0.6	0.7
	C	28.1	36.4	30.6	2.4	1.0	1.0	0.6
	D	27.1	35.9	31.6	2.7	0.9	1.1	0.6
	E	28.0	34.6	32.2	3.0	1.0	0.8	0.4
V	A	25.2	34.1	34.9	2.6	1.8	0.7	0.6
	B	28.6	34.5	31.6	2.7	1.6	0.6	0.4
	C	28.3	33.0	32.8	3.7	1.0	0.7	0.4
	D	26.3	36.1	32.3	2.9	1.1	0.8	0.5
	E	28.6	34.6	31.9	2.7	1.2	0.5	0.5

lines 2 and 4 of Table 3 are better estimates of it than either set of observed values, for they depend only on the grand mean for each constituent, and the error of this grand mean is small. Each operator's analyses should be approximately normally distributed around his own mean with σ equal to the appropriate entry in line 2 or 4 of Table 3. The frequencies of differences less than σ, intermediate between σ and 2σ, and greater than 2σ for the entire suite of analyses ought to be about in the ratio .65:.30:.05. For the set of 25 analyses of each constituent, there should be about 16 differences $<\sigma$, 7 or 8 in which $\sigma<\Delta<2\sigma$, and 1 or 2 which are larger than 2σ. The observed differences for the two runs are calibrated in this fashion in Fig. 1 (the hundredths place was retained in the theoretical errors for this comparison); for each constituent the bar on the left gives the frequency of $\Delta<\sigma$, the center bar shows the frequency of $\sigma<\Delta<2\sigma$, and the one on the right shows the frequency of $\Delta>2\sigma$. In such small samples an exact check is not possible and even a close check is rather encouraging. With the exception of quartz differences, which are about the same in both runs, the second run shows considerable improvement over the first for essential minerals. It is rather curious to find that *least* difficulty was encountered with accessory minerals.

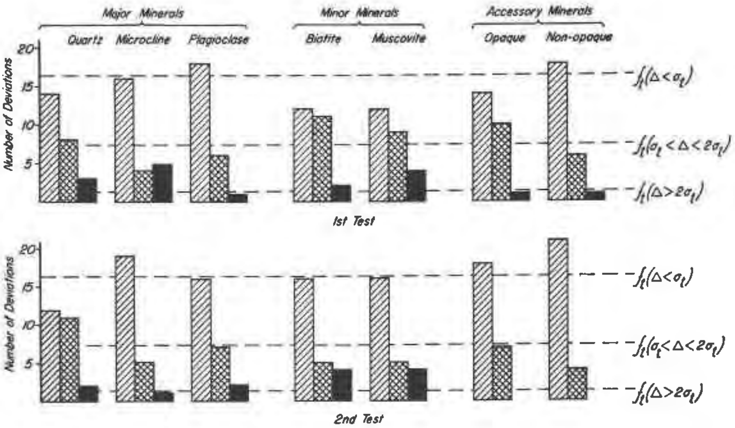

FIG. 1. Deviations of individual results from operator means. (For each constituent the bar at the left shows the number of deviations $<\sigma_t$, the center bar shows the frequency of deviations such that $\sigma_t<\Delta<2\sigma_t$, and the bar at the right shows the frequency of deviations $>2\sigma_t$. The theoretical [normal] frequencies are shown by the dashed lines.)

The full array of mean squares generated by the second run, calculated from the data of Table 4*B* and used in the variance analysis, is shown in Table 5. The discussion of the preceding paragraphs takes care of the last two columns of the table so that only analyst and slide differences[2] remain to be noted.

No highly significant differences between slides have been established even though each slide has been analyzed five times. The slides are no doubt different, but the differences between them must be very small. There is some indication of small differences in quartz and plagioclase, but in both cases the F ratios are well below the .01 point.

Perhaps the most important result of the test is contained in the first column of Table 5. There seem to be no significant differences between analysts so far as plagioclase, microcline, and opaque accessories are concerned, and for quartz the F ratio is barely beyond the .05 level. If this F ratio is not itself in error, e.g., if it is not indicating a difference when none in fact exists, the variance components of quartz attributable to interanalyst differences may be estimated as 2.0455/5. Expressed as a standard deviation this comes to $\sqrt{2.0455/5}$ or 0.6 per cent of the whole. As the mean quartz content is 27.2, the error component attributable solely to operator differences amounts to 2.2 per cent on a relative basis. The estimation of biotite by the group is probably still affected by incomplete standardization of tabulation procedures; at any rate, analyst difference is highly significant for biotite and nonopaques taken

Table 5. Mean Squares Calculated from Table 4*B*

Mineral	Source of Variation			(Theoretical Error)^2
	Analysts	Slides	Error (Obs.)	
Quartz	2.7884*	3.3594*	0.8949	0.7921
Microcline	1.6316	0.4436	1.1204	0.9125
Plagioclase	1.2846	3.2476*	0.8164	0.8723
Biotite	1.6694**	0.6074	0.3259**	0.1224
Muscovite	0.1124	0.1784	0.1232**	0.0466
Opaques	0.0496*	0.0546*	0.0141	0.0298
Nonopaques	0.1246**	0.0226	0.0106	0.0155

* See footnote 2.

---

[2] The object of the calculations—actually the object of the test—is to obtain estimates of variance which may be compared with each other. Identifying sources of variation by subscripts, the ratio $F=V_a/V_e$ is a measure of the differences between analysts in terms of analytical error. Use of the F ratio and a tabulation of the .05 and .01 significance levels are given in many modern elementary texts (e.g., Snedecor, 1946, pp. 220–225 and accompanying discussion). A sample mean square yielding an F value exceeding the .05 point is customarily denoted by a single asterisk. One giving an F value exceeding the .01 point is denoted by a double asterisk. These conventions are followed in Table 5.

separately but only suggestive for the sum (biotite+nonopaques). Some of the analysts may be recording chlorite which replaces biotite partly under biotite and partly under nonopaques.

At the literal level the experiment may be dismissed as showing only that there are at MIT five petrographers who get about the same results as one petrographer at the Geophysical Laboratory. The design and execution of the test, however, justify more general inference. Viewing the results in the most favorable light, we may conclude that with reasonable care and practice there is no reason why the reproducibility of quantitative thin-section analysis should not easily equal that of analytical methods ordinarily regarded as inherently far more precise. The common practice of calculating norms or modes from chemical analyses is far more expensive and leads to results no more precise than those of thin-section analysis, providing, of course, that the material in question is neither glassy nor excessively coarse.

REFERENCES

CHAYES, F. (1949), A simple point counter for thin-section analysis: *Am. Mineral.*, **34**, 1–11.

SNEDECOR, G. W. (1946), Statistical Methods, Iowa State College Press, Ames, Iowa.

TECHNICAL NOTE

A study such as this offers nothing novel to the professional statistician, and in order to keep the discussion within bounds no attempt has been made to satisfy readers uninterested in or completely unfamiliar with statistical analysis. For this latter group the argument is necessarily one of example rather than precept; it will have been successful if it succeeds in whetting their appetites. Excellent descriptions of the experimental design and calculations will be found in Snedecor (1946, pp. 265–274) as well as in most other modern elementary texts.

Readers familiar with practical statistical work will have noted from Table 1 that the original test was set up in a Latin square. It was anticipated that significant operator biases would be found, but that the analysts would have sufficient opportunity for practice in advance of the test so that their improvement after its start would not be marked enough to influence the result. A significant mean square for order would thus imply instability which could not be attributed to learning or incompetence but would have to be regarded as one of the hazards of the method. The expected biases did not emerge in the first test; as has already been noted, however, this might have been because analytical error variances for essential minerals were all larger than anticipated. Although operator differences in the first run were not significantly larger than observed error variances, several of them *were* larger than the appropriate theoretical error variances, and a more extensive test had already provided confirmation of the theoretical values.

Maximum interest in the second run therefore focused on the analytical error. In the first run the effect of "order" was negligible, and when this proved to be the case in the second also this criterion was abandoned. Its four degrees of freedom were pooled with the twelve allotted to "discrepance" in the original square, thus materially strengthening the test as a measure of analytical error. With elimination of one factor the Latin square degenerates into a randomized block design, and it is in this form that calculations from the final data are shown in Table 5.

The fact that the excess of analytical error noted for major constituents in the first run is not in evidence in the second implies that the basic assumptions required for proper application of the design were not valid when the test was begun. The simplest explanation is that the practice period preceding the test was not adequate and that some of the operators improved notably during the first run. This would find expression as an analyst-order interaction, and, in either the square or block design, variation occasioned by this interaction would enlarge the observed error variance. The plausibility of this explanation is considerably increased when it is recalled that of the four possible interactions the other three all involve the sections, and major mineral differences between slides have not been clearly established by either run. Interactions involving the slides should therefore all be negligible.

Manuscript received March 12, 1951.