Evaluation of Chemical Analyses on Two Rocks

W. J. YOUDEN

National Bureau of Standards

Interlaboratory and round robin test programs can provide information, not only on a test procedure under investigation, but on the participating laboratories as well. A simple graphical technique is proposed to aid in the comparisons between laboratories.

INTRODUCTION

Test methods in industry are often submitted to an inter-laboratory or round robin test program. The avowed intention is to find out how satisfactory the test procedure is in actual practice. The point of view may be taken that the participating laboratories are just as much under study as the test method. If eight or nine laboratories out of ten get good results, perhaps attention should be directed to the laboratories having difficulty instead of to the test method. A published example of an extensive and carefully controlled study of the chemical analysis of rocks is reviewed in this paper with particular consideration given to the participating laboratories.

Geological Survey Bulletin 980, "A Cooperative Investigation of Precision and Accuracy in Chemical, Spectrochemical and Modal Analysis of Silicate Rocks" 1951, reports chemical analyses by 34 cooperating laboratories on a granite rock and a dibase rock. The analytical results from the various laboratories posed a difficult problem of evaluation. The report discusses the use of the modal value, the use of the mean of all of the results, and the use of a mean based upon a consensus of more or less closely clustered values. The report includes scatter diagrams of the results for each determination. A separate scatter diagram was made for each rock. The selection of the results to be included in any consensus cluster is not an easy one. This note approaches the problem by an extension of the familiar scatter diagram. The purpose is to pick a subset of the laboratories and use the results from these laboratories for evaluating the analytical results.

ANALYSIS OF DATA

Thirty of the laboratories reported on both rocks. These data are recorded in Tables 1 and 2 of the Geological Survey Bulletin 980. Each of these thirty laboratories makes available a *pair* of results; one for each rock. Consider the SiO_2 determinations. Plot these paired values using the x axis for the granite results and the y axis for the dibase results. Each laboratory provides a point in a two dimensional scatter diagram shown in Figure 1. In theory, for statistically independent measurements with no systematic differences between laboratories,

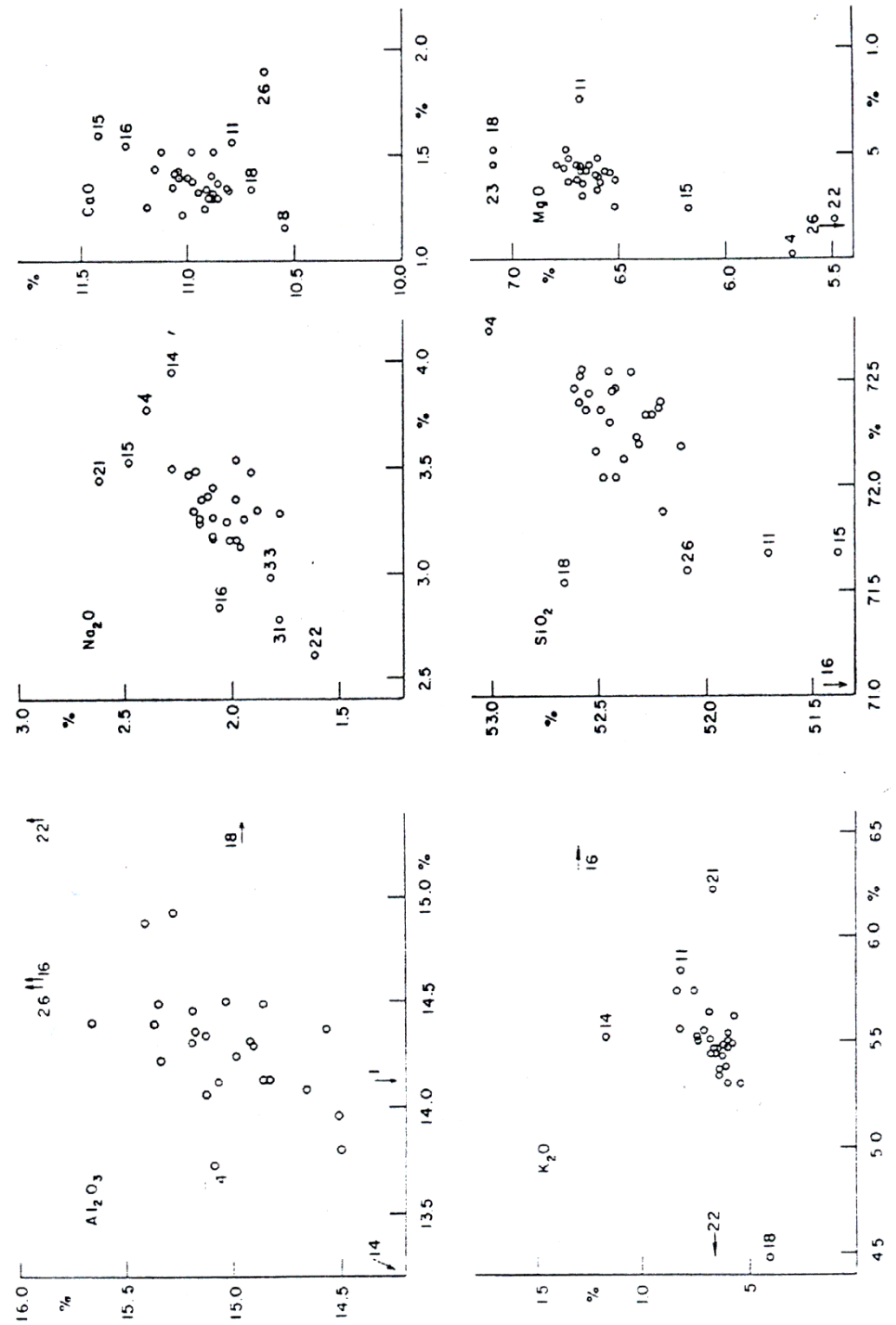

FIGURE 1. Dibase rock results (y-axis) plotted against results obtained on granite rock (x-axis The numbers in the graphs identify the outlying laboratories.

TABLE 1. *Tabulation of laboratories with very extreme results[a] for six of the elements.*

Element	Laboratory Number													
	16	18	4	15	22	26	11	14	21	1	S	23	31	33
Si	*	*	*	*		*	*							
Al	*	*	*		*	*		*		*				
Mg		*	*	*	*	*	*					*		
Ca	*	*		*		*					*			
Na	*		*	*	*			*	*				*	*
K	*	*			*		*	*	*					

[a] The asterisk denotes very extreme results. The 16 laboratories not listed had no extreme results for these six elements.

the points should be scattered in a circular pattern centered on the center of gravity of the whole group. Almost always the pattern of points departs from theory in two noticeable ways. First, there is a main body of points, more or less elliptical in shape with the long axis of the ellipse inclined at an angle of approximately 45° to the x-axis. Second, a small minority of the points are individually separate from the main cluster and, in most cases, these points tend to lie near the extended long axis of the ellipse.

Both these departures from theory have a reasonable explanation. The center of gravity of the main cluster represents the consensus of the laboratories for the method employed in the determination. The coordinates of this centroid are probably the best estimates of the unknown values that can be made from the data. The displacement of this centroid, from the true values for the two rocks, is largely along a 45° line through the unknown true value because any systematic error in the procedure may be considered to apply (in most cases) equally to both rocks. If the method, as used, tends to give high results by P percent, it presumably does this for both materials and thus the x and y displacements are equal and the point is displaced along a 45° line. Now any one laboratory is apt to have its own systematic error *relative* to the systematic error in the procedure itself. For the moment forget about the true values and focus on the centroid which represents a working datum of reference. A laboratory which tends to get results higher than the other laboratories will be displaced from the centroid upward along this 45° line because it gives the same positive bias to each material. A laboratory with an excessively large relative bias will be considerably displaced along the 45° line. The general elliptical shape results from the collection of these individual biases.

These two dimensional scatter diagrams are useful in the evaluation of the results. There is a reinforcement of the judgment in setting aside certain determinations when the scatter diagram shows an emphatic discrepancy by one laboratory for both rocks. There are a few instances where the displacement from the main group is accounted for by one of the pair of results and one cannot help but wonder if some simple slip in calculating or typing is the explanation.

In any event an examination of the scatter diagrams for SiO_2, CaO, K_2O, Na_2O, MgO and Al_2O_3, shown in Figure 1, shows approximately six points in each diagram well removed from the main group. The two dimensional departure of the points from the centroid gives a good basis for suspecting these out of line results.

The foregoing does not exhaust the inferences to be drawn from these scatter diagrams. The elliptical character of the cluster has been pointed out. The more vulnerable the analytical procedure is to individual laboratory bias, the greater the tendency of the main cluster to assume an elliptical pattern. By this method of plotting a visual comparative appraisal of different procedures (for the same or different elements) may be made. Procedures that give a generally circular cluster indicate that, should a systematic error be present in the procedure, all the laboratories are afflicted with this same systematic error both in direction and magnitude or that the systematic errors are small with respect to the precision of the method.

TABLE 2. *Tabulation of Statistics Calculated From the Data.*

Element and rock		Table 20, Bull. 980		Accepted laboratories[a]			F[c]
		Av. of all	Av. of consensus	Av.	S.D.[b]	$S.D._{a_p}$	
SiO_2	G	72.22	72.45	72.31	.127	.028	2.40
	D	52.25	52.45	52.41			
Al_2O_3	G	14.44	14.30	14.31	.161	.036	5.30
	D	15.23	15.10	15.08			
MgO	G	0.39	.0.45	0.41	.049	.011	3.28
	D	6.52	6.65	6.65			
CaO	G	1.42	1.35	1.36	.090	.020	1.12
	D	10.95	10.95	10.97			
Na_2O	G	3.26	3.35	3.32	.097	.023	2.17
	D	2.05	2.10	2.06			
K_2O	G	5.51	5.45	5.48	.063	.014	3.18
	D	0.71	0.65	0.65			

[a] Rejection of all the data from laboratories 16, 18, 4, 15, 22, 26, 11, 14 and 21. Also the determinations with an asterisk for laboratories 1, 8, 23, 31, and 33 in Table 1.

[b] The precision component, i.e., the standard deviation of a single determination. The next column gives the standard deviation for the average of the accepted laboratories. These standard deviations are based on an estimate of the precision.

[c] F is the ratio of the square of the standard deviations for laboratories and precision. A ratio close to unity is expected if the laboratories all have the same systematic error. Values of F larger than about two are considered good evidence that the laboratories have individual systematic errors that differ from laboratory to laboratory. An F value of 3 indicates a systematic error component about equal to the precision component.

The collection of scatter diagrams makes possible a reinforcement of the rejection procedure. The laboratories identified with the outlying points for the determinations of SiO_2, Al_2O_3, MgO, CaO, Na_2O, and K_2O are shown in Table 1. These points are also identified in Figure 1. All graphs use the same unit space for one percent to facilitate comparison of the procedures.

Certain laboratories are very much in evidence as the source of the outlying points in these six diagrams. In fact six of the 30 laboratories account for 25 of the 39 outlying points. The average number of outlying points for all 6 tests combined is about 1.3 per laboratory. If the outlying location is only a matter of chance, we may use the Poisson distribution to calculate the expected proportion of laboratories with none, one, two, or more outlying points. We could expect just about one of the 30 laboratories to be out in four or more of the six diagrams. Instead, six laboratories achieve this unenviable distinction. Two of these laboratories are remote in five of the diagrams. We would also expect only eight or nine laboratories never to be out of the pattern if this event depended on chance. It is encouraging to find sixteen of the thirty laboratories avoid the outfield. Five of the laboratories turn up in the outfield just once, and in two of these cases just one of the rock determinations is responsible. Except for the particular determinations in question, the work of these five laboratories appears appropriate for retention. This leaves 21 or 20 laboratories for determining a consensus both as to mean and dispersion for the analytical procedures.

Scatter diagrams for Fe_2O_3 and FeO were also prepared. The one for Fe_2O_3 showed that considerable difficulty was experienced with this determination because about half of the points showed extremely large departures from the residual compact group. Almost all the laboratories that had trouble with other determinations obtained poor results with the Fe_2O_3 determination. Similarly the eight outlying points for the FeO results all came from laboratories having trouble with three or more of the six determinations shown in Figure 1. The FeO data present a special problem because the dibase results show a much greater dispersion than the results for the granite.

Random and Systematic Errors

The two dimensional scatter diagrams illustrate an important partition of the errors of the reported results. There are first the purely random errors which are revealed by the scatter of repeated determinations made on the same rock within a given laboratory and second the individual laboratory systematic errors measured from the consensus of the results from the qualifying laboratories. Earlier it was pointed out that if there were no differences in the systematic errors, the pattern of points would tend to be circular. On the other hand, if the individual laboratories had perfect precision, i.e., in this case could check themselves exactly to 0.01 percent, the scatter would arise solely from the individual systematic errors. Indeed, in this improbable state of affairs, the points would lie exactly on the 45° line through the centroid. Inasmuch as both types of error are present, the points depart more or less from the 45° line but retain this line as the major axis for the scatter of the points.

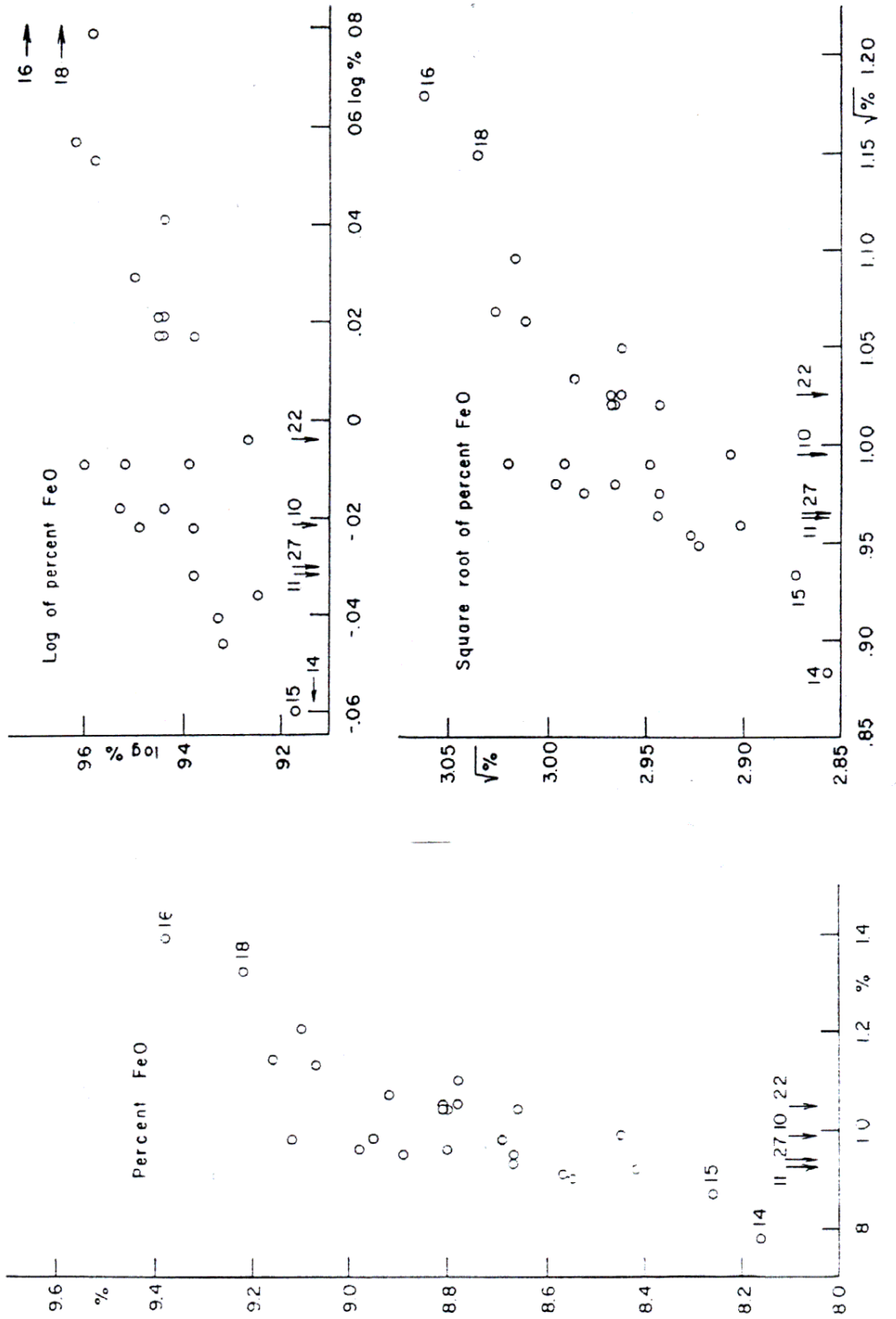

FIGURE 2. Three ways of plotting results for ferrous oxide on dibase rock (y-axis) and granite rock (x-axis). The numbers in the graphs identify the outlying laboratories.

The departures of the points from the 45° line, as measured by the lengths of the perpendiculars from the points to the 45° line, correspond to the precision component of error. The lengths of the segments along the 45° line from the centroid to the feet of these perpendiculars measure the *relative* systematic errors. It is instructive to square each length *along* the line and sum these squares. Also sum the squares of all the perpendicular lengths. The combined total of these two sums will be found equal to the sum of the squares of the deviations from the mean for the granite added to a similar sum for the dibase. The sum of the squares of the deviations from the mean is a familiar step in calculating a measure of the dispersion of a set of results. Given just one set of results, say those for granite, no separation between precision and systematic components of errors is possible. When the results on two materials are available, the position of a point representing the two results may be represented by its distance (from the centroid) *along* the 45 degree line and by its perpendicular distance *from* the line. These two distances are identified with the relative systematic error and the precision error respectively. Given results on two materials, the dispersion estimates can be expressed in this alternative and more revealing form. Some readers may wish to consult a statistical text on the analysis of variance of data with a two way classification.

It must be pointed out that when rocks of sharply different composition are involved (either as to amounts of the elements or the presence of possibly troublesome elements) the precision component will also include any change in the systematic errors for the laboratory. If the two materials are rather similar, there is every reason to expect the systematic error to be the same for both materials in any one laboratory. When the *range* of the scatter of the results is about the same for both rocks, this complication is not likely to be present to any important degree. Another complication is the possibility that the precision error depends upon the amount of element present. Thus if there is a ten-fold difference in amount of element present in the two rocks, the scatter along one axis may be visibly greater along one axis than along the other axis and this in turn will cause a departure of the major axis of the ellipse from the expected 45°. In this event, it is often helpful to convert the data to logarithms before plotting or undertaking a statistical analysis. This statistical device will give equality of variance under the assumption of a constant coefficient of variation. The FeO determinations afford an excellent illustration of this complication.

In Bulletin 980 there is a notation that some of the results listed are based on duplicate determinations. No allowance was made for these duplicates because experience indicates that the systematic errors dominate the precision component. The effect is to bring these points a little closer to the 45° line than single determinations would fall. In consequence the estimate of the precision has been slightly favored. The other assumption made is that of a common precision for all the accepted laboratories. This is a reasonable approximation because the differences in precision are minor compared with differences in systematic error and, in any event, many determinations would be required to establish that such differences in precision existed. Observe also that many of the points far removed from the centroid are reasonably close to the 45° line

indicating that even the presence of large systematic errors is not necessarily tied in with precision. The estimate of precision, using the data from the selected laboratories, provides a basis for predicting the extent of agreement that can be expected between laboratories that in one way or another adjust their procedures by reference to a standard sample.

The scatter diagram first prepared for FeO showed a very much greater spread for the dibase results than for the granite rock. (Fig. 2). Consequently the logarithms of the data were used to prepare a graph. After this transformation there was a greater spread in the granite values than for the dibase values. There are two rather well defined clusters in this diagram and this suggests the important clue that there were two different analytical procedures. The centroids for the two clusters differ mainly in the means for the granite rock suggesting that it is on this rock that the two analytical procedures disagree. Finally the data were plotted using the square roots of the data. The axis of the major cluster is now much closer to making the expected 45° angle with the x axis. The suggestion of two distinct groups in the pattern continues in the diagram based on the square roots. Statistical estimates based on composites of two differing groups have no chemical utility. The examination of these data puts the problem of FeO determination right back in the laboratory.

The two dimensional scatter diagram shows the need for a careful consideration of the method for determining FeO. The 45° line through the centroid of the data should run through the mass of points in such a way that a ruler pushed along perpendicular to the 45° line should encounter points on the left and right of the line in a random order. Any runs of points on one side, succeeded by long runs on the other side are a signal for caution. Statements about the procedure that hold only for a particular rock are not helpful. Bulletin 980 did segregate the data on the basis of the procedure employed wherever this information was available. Not much can be learned from a result unless it is accompanied by information on the method used.

In Table 2 there is shown the *precision* component for the determinations of SiO_2 , Al_2O_3 , MgO, CaO, Na_2O, and K_2O. This standard deviation shows the optimum performance that could be expected for these determinations should the laboratories achieve the same systematic error. As a matter of fact, these very rocks, if the compositions become accepted, are intended to serve as standards and would, by providing a standard reference point, achieve this goal of a common systematic error. The amount of this common systematic error is, of course, exactly the amount by which the designated composition departs from the unknown true composition.

Besides the precision component there are listed the averages of the results from the selected two thirds of all the laboratories. It is interesting to compare this column with both the averages for all laboratories and the consensus that have been taken from Table 20 in Bulletin 980. The precision component does not contain the component of error arising from variation in the systematic error among the laboratories. A measure of this source of error is provided by the column of F values. Large values of F show a real need for standardization among the laboratories.

In Bulletin 980 emphasis is placed (p. 12) upon the possibility that some extreme value may in fact be closer to the true value than an average based upon all the data, a consensus of selected results, or some modal value. This invites the reply that whatever comfort this may be to a laboratory with a point at one end of the 45° line, the same gesture adds to the discomforture of those laboratories on the opposite end by making them even further off from the true value. No one will seriously contend that the centroid is without bias. But in a practical world, men must adhere to some standard, even a standard made by men. The laboratory with an extreme position along the 45 degree line *may* be the most nearly correct laboratory and the other laboratories all afflicted with a systematic error of the same sign, but evidence is required to support such a position. Unless, and until, such evidence is forthcoming, the use of a consensus value will persist. The double sample plotting scheme described in this paper should prove useful in selecting those laboratories doing consistent work near the consensus.

Literature Cited

(1) Youden, W. J., *Industrial Quality Control*, Vol. XV, No. 11, May 1959.
(2) Youden, W. J., *Ind and Eng. Chem.*, Vol. 50, page 83A, August; Page 91A, October; Page 77A, December 1958.