DISCUSSION OF PAPERS ON FULL-SCALE TESTS FOR ENGINE OILS

MR. R. K. WILLIAMS[1] (*presented in written form*).—The authors are to be congratulated for developing the test procedures and for their diligent efforts to improve them. The five sequence tests are an improvement over their predecessors—the L-4, the FL-2, the EX-3, and the LS-5 tests—and, when properly conducted and interpreted, they should provide a sound basis for developing better motor oils. That they now constitute the basis of nearly every car manufacturer's specification for crankcase lubricants is testimony to their value and importance.

Because the tests are both widely used and expensive to conduct, we are pleased at the efforts that are being made by ASTM to refine them and thereby improve their repeatability. This is an area where continued efforts will be necessary if the cost of developing and approving oils on the basis of these tests is to be kept within reasonable limits. Since the cost ranges from $1700 to $1900 per test, we can afford to expend considerable time and money toward improving the repeatability and reproducibility of the results obtained and it is on this aspect that we wish to focus our discussion.

Repeatability is defined as the degree to which a given laboratory can obtain the same answer in tests which are considered to be conducted under identical conditions. Reproducibility is used to describe constancy of answers from different laboratories. When the originator describes the conditions under which a test is to be conducted, he attempts to specify within narrow limits all of the conditions which he thinks will affect test results. Frequently, the user of the tests naively assumes that all he must do to get the correct answers is to follow the test instructions as outlined. Unfortunately, very few of us are perceptive enough to foresee all of the important test variables, and it is only through costly experience that we finally learn of them. We do not yet know all of the variables that must be controlled to achieve the desired repeatability in any of the sequence tests.

Mr. Potter and his associates have contributed materially to an improvement in the sequence V tests by setting up a system for providing a common gasoline, certified as to severity on the basis of tests in several laboratories. The need for such a system is verified by the data in Fig. 1. Though both of the fuels met the written specifications for the MS-06 gasoline then specified for sequence V testing, the results with fuel A were appreciably less severe than with fuel B. Obviously, any comparisons between oils could be obscured by the large differences between these fuels. The Ford engineers should be commended for eliminating such a possibility through their present fuel certification program.

While the CRC has prepared excellent instructions for rating sludge and varnish, there remained the possibility that differences in interpretation and rating techniques were contributing to variability of test results between

[1] Lubrizol Corp., Cleveland, Ohio.

Copyright© 1963 by ASTM International www.astm.org

laboratories. To explore this possibility, raters from four different laboratories were asked to rate, in duplicate, parts from tests on three oils—one of good, one of intermediate, and one of poor sequence V quality. All inspectors were asked to use their regular rating techniques, after which results were reviewed with an eye toward eliminating the most glaring differences. Results are shown in Fig. 2, which shows the 95 per cent confidence limits for the sludge and varnish ratings obtained. The results are expressed as the number of rating units within which 95 per cent of the ratings for the *same part* would fall when rated by different raters. For the individual parts, the raters usually could be expected to agree within two numbers using a basis of ten equals clean. The greatest sludge variability was in rating of the oil screen. This was later found to be due to differences in proportion of the total area attributed to the sides and bottom of the screen. It was agreed, as a result, that the sides of the screen should be considered 60 per cent of the total area in all future work.

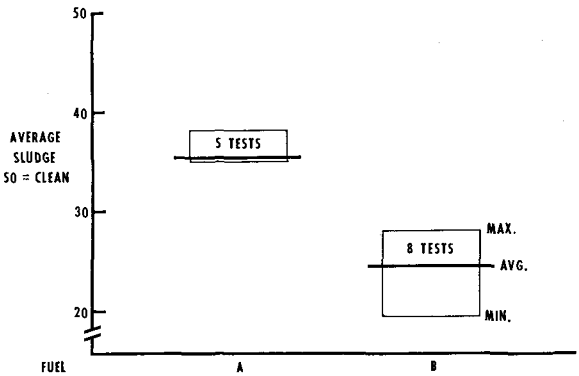

FIG. 1.—Lincoln Sequence V Sludge Ratings Using REO 132 Reference Oil.

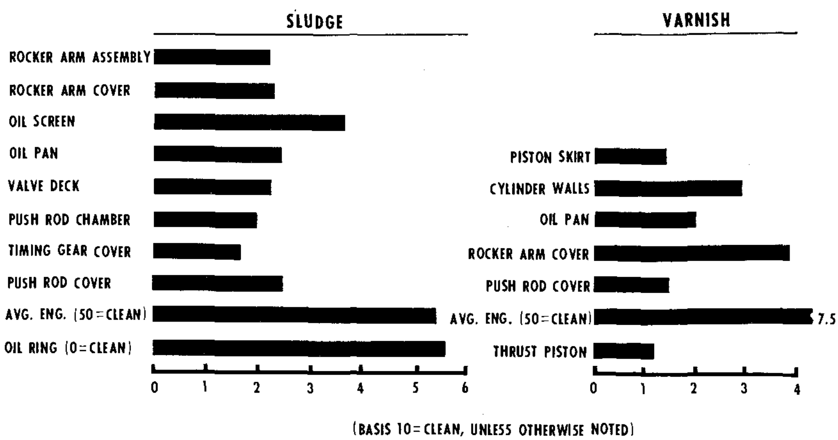

FIG. 2.—Lincoln Sequence V Ratings: 95 Per Cent Confidence Interval. Same part rated by different raters.

Similarly, the largest differences in varnish ratings were found in the rocker arm covers and cylinder walls. The former were resolved by agreeing to rate one cover for varnish by removing all of the sludge in a uniform manner. Cylinder wall varnish variations will be reduced by resorting to a standard lighting system in all future rating of these parts.

One other point should be mentioned in connection with Fig. 2. If a group of experienced raters were to rate the *same engine* prior to the resolution of some of the above differences, the over-all sludge ratings would be expected to vary by about five units on a fifty equals clean basis. This variability is exclusive of any variations in test techniques among laboratories. It is important to recognize this variable in interpreting results from different laboratories.

A similar program was conducted in connection with the rust rating of parts from sequence I to III Tests. Results are shown in Fig. 3. Among the individual parts the largest rating variation occurred for cylinder bores, wrist pins, and hydraulic valve lifter balls. It is expected that the cylinder bore rust rating variations will be reduced through universal adoption of the K-H 11-in., 8-w fluorescent safety lamp. Over-all lifter and engine rust ratings for a given engine by different raters should agree within $1\frac{1}{2}$ rating units. While this variation is appreciable and should be reduced, it is believed to be only a small part of the variations due to other causes that have been inherent in the sequence I to III Tests.

In any event, differences in rating for the same parts could probably be kept to a minimum through a regular cooperative exchange of ratings. Such a program would resemble the ASTM national exchange program for antiknock rating of gasolines and is a logical area for ASTM activity in the future.

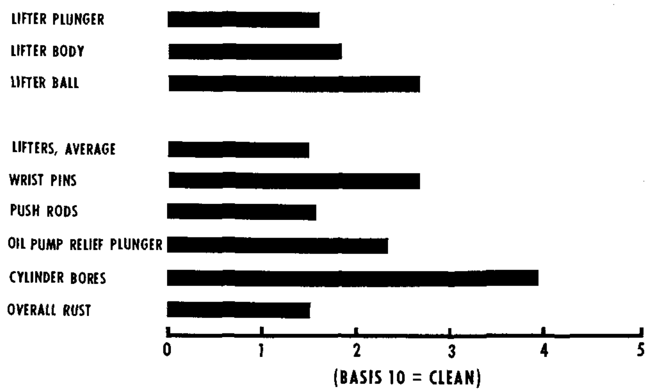

FIG. 3.—Oldsmobile Sequence I to III Rust Ratings: 95 Per Cent Confidence Interval. Same part rated by different raters.

Much of the Bennett and Kabel paper is devoted to a discussion of reproducibility and those factors necessary to reduce unwanted variation in test results. We wish to commend the authors for their emphasis on this, since we have certainly had our problems in achieving an acceptable degree of repeatability and severity with respect to rust; nor are we consoled by the fact that we are not alone in this respect. Part of the problem stems from the fact that we do not know what the repeatability and reproducibility of the test really are, and if we did, we would still not know how to define the words "Satisfactory" and "Unsatisfactory" as used in the authors' Tables II and III. At a cost of $1900 per test, and using the authors' estimate of 1200 tests per year by the petroleum industry, we are spending over two million dollars per year on these tests alone. A great percentage of these tests have been made with the General Motors Reference Oils, and much of the expense could be saved or better spent in developing improved motor oil formulations if the factors responsible for the wide variability in test results could be discovered and eliminated, thereby reducing the number of repeat tests required to establish valid answers as to product quality.

Bennett and Kabel repeatedly emphasize the importance of adhering to details of operating procedure, engine and accessory buildup, and maintenance. In our laboratory we have always endeavored to adhere to the "rules and regulations," as specified by the originators of this test; yet we have encountered unexplained difficulties. Figure 4 will illustrate our problem. Results are shown for consecutive runs in four different engines, all tested with GMR 6063 borderline rust reference oil. Results are reported for three consecutive tests on two engines with inlet air humidity at 80 grains and four consecutive runs at 100 grains. All engines were new at the beginning of each series, and engines were conditioned according to General Motors' instructions prior to each test. Except for the use of the 100-grain humidity level in some of the tests, other operating conditions were in all respects within the limits specified. The points connected with the solid lines show a fairly definite trend toward increasing rust severity of the engine with increased run number. Though present, this trend is not so pronounced for the lifters as for the engine. It is obvious that the unknown engine condition effects responsible for these trends could easily obscure

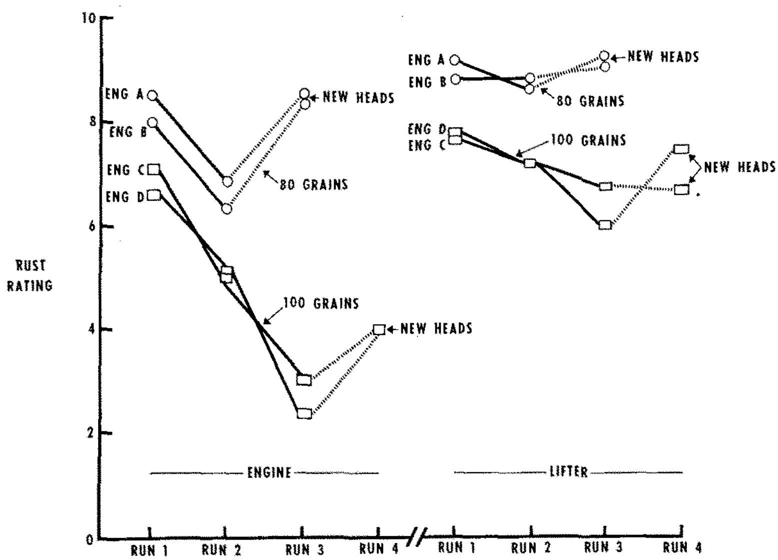

FIG. 4.—Effect of Engine Condition and Humidity on Rust Using GMR 6063 Reference Oil.

the effects of variations of humidity within the 80 to 100-grain range.

The dotted lines indicate the effect of installing new cylinder heads even though this change was not deemed necessary on the basis of careful measurement of valve-to-guide clearances. As can be seen, test severity was reduced with the new heads. Our efforts to correlate the effects of the head replacements on rust with measurements of exhaust gas in the crankcase have not been successful, although we believe that the change in severity must be associated with leakage of engine exhaust past the guides and into the crankcase atmosphere. This may partially explain the authors' current recommendation that new cylinder heads be used for each test.

Figure 5 illustrates blowby measurements observed in our laboratory for a number of tests carried out according to specification. We agree that blowby and percentage of burned exhaust in the crankcase are important variables and we attempt to control them by rigid adherence to the ring gapping and bore conditioning specified by the authors. Figure 5 suggests that present limits may be somewhat unrealistic in terms of today's engines since it is virtually impossible for us to maintain blowby for sequence II and sequence III Tests, both within limits by these methods. Since we are mainly interested in engine rusting, we attempt to hold blowby in Sequence II well within limits despite the fact that this results in blowby being slightly low in Sequence III. We would appreciate some relief from the authors in this area.

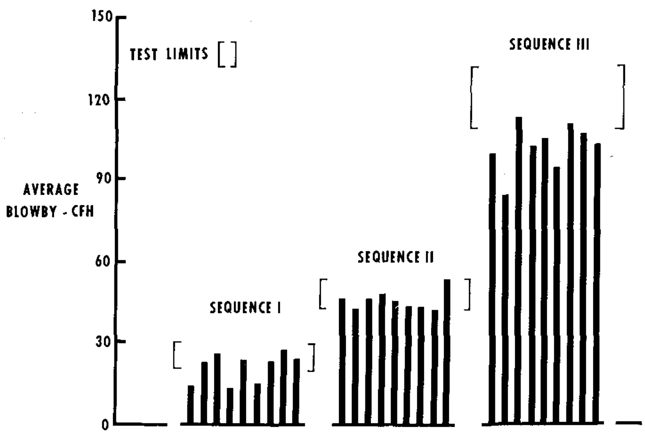

FIG. 5.—Oldsmobile Sequence I to III Tests: Average Blowby Rates for Nine Tests.

The above are only a few of the many unanswered problems in connection with the sequence I to III Tests. There are many more. Unfortunately, all laboratories differ both as to severity and to repeatability. Many have no clear-cut idea where they stand on either score. All would like to know, and most would like to be in a position to adjust their severity to a common level. Unless some practical means of doing so can be suggested, much of the practical value of the MS Sequence Tests will be lost. Some way must be provided for the conscientious tester who follows all the rules and still gets the wrong answers, to bring himself into line. We believe Fig. 6 offers a possibility in this respect.

Finally, we would like to suggest the initiation of an exchange program similar to the ASTM National Motor Fuel Exchange Program which would enable

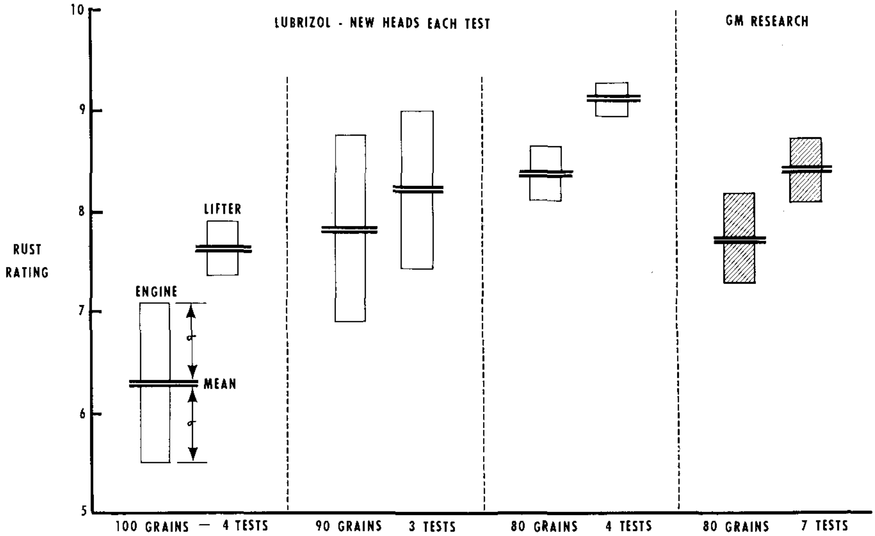

FIG. 6.—Oldsmobile Sequence I to III Tests: Effect of Humidity on Rust Using GMR 6063 Reference Oil.

Here results are shown for tests using GMR 6063 reference oil at three different humidities. The heavy horizontal lines are averages; the standard deviations for the groups of tests are also shown. New cylinder heads were used in each test, and in no case had more than two tests been conducted with the engine prior to the run for any of the Lubrizol tests. The mean values for over-all engine and tappet ratings confirm the increase in rust severity with increasing humidity previously reported by the authors. They also show that a greater spread in ratings will occur for the more severe conditions. For those laboratories adhering to the written procedures in all respects, minor variations in humidity control offer a simple solution to the severity problem, and we would suggest that this variation be permitted.

each laboratory to determine severity and reproducibility with respect to other laboratories in a manner which would eliminate all possibility of bias in the results. This could be accomplished by testing unknown oil samples at periodic intervals and circulating the results. We feel this procedure with respect to all of the sequences described in this symposium would make the tests more meaningful and would help laboratories to reduce variations in results to a minimum.

MESSRS. P. A. BENNETT AND R. H. KABEL (*authors*).—Mr. Williams has implied that test procedures should not be used until all of the important associated test variables have been identified. At first thought this would seem to be a necessary prerequisite. However, almost everyone involved in engine testing will agree that this idealized situation has never existed. We do not know of a single engine test procedure that would meet this requirement.

In addition, the matter of rust testing is probably the most complicated of any engine test procedure. As often pointed out, all of the important test variables associated with rusting tendencies have not been identified.

If repeatable results are to be obtained, it is imperative that all controllable parameters be maintained essentially identical from one test to another. This means that all of the steps detailed in the written procedure must be conscientiously followed. In addition, all other aspects of engine rebuild and operation, including those not specifically controlled or outlined in the written procedure, must be handled in the same way each time. When a variation exists, even though its importance may not be presently recognized, significant changes in test results can and do occur. Most repeatability problems have resulted from inadequate control of test variables.

TABLE I.—EFFECT OF CYLINDER WALL FINISH ON BLOWBY RATE AND OIL CONSUMPTION

Cylinder Wall Finish μ in.	Ring Gap, in	Blowby Rate, cu ft per min			Oil Consumption, qt
		Sequence II	Sequence III	Sequence III − II	
20	0.022	0.81	1.72	0.91	3.2
23	0.022	0.78	1.66	0.88	3.3
22	0.022	0.61	1.45	0.84	5.0
30	0.022	0.78	2.03	1.25	2.5
35	0.23	0.80	1.83	1.03	2.4
46	0.022	0.79	1.84	1.05	9.2

Poor repeatability can also result from attempts to use engine parts for too many tests. This is usually poor economy. It generally results in the introduction of complications which can influence the results, as shown in Mr. Williams' Fig. 4.

With regard to the variation in rust ratings of different parts, it should be emphasized that the rusting of some engine parts is more sensitive to small differences in engine operating conditions than others. Much more consistent answers are obtained by using average ratings. Therefore, the "average lifter" and "average engine" rust ratings should be used in making comparisons based on sequence I, II, and III data.

It should also be pointed out that the variation in ratings indicated by Mr. Williams' Fig. 3 is not as extreme as it first appears. One of the raters involved in this study had never rated parts for rust before; a second had only limited experience; and a third used no standards, but relied entirely on memory in making the rust ratings. These factors would result in an increased spread in the ratings. At the same time, such comparisons point out the need for careful coordination of rating procedures if reproducibility problems are to be minimized.

Failure to reproduce the blowby limits specified for sequence II or I, II, and III tests is usually related to cylinder wall finishes. As indicated in the accompanying Table I, using smooth wall finishes results in a relatively small difference in blowby rates between sequence I and sequence III. The recommended difference in blowby rate is observed with cylinder walls having a 30 to 35 μin. finish. If wall finishes are excessively rough, high oil consumption rates are observed.