Use of Marshall Specimens in a Nuclear Asphalt Content Gauge

JAMES L. BURATI, JR., THOMAS E. FREEMAN, AND HOKE S. HILL, JR.

An evaluation of the Troxler 3241 asphalt content gauge with regard to its capacity to consistently measure asphalt content of Marshall specimens is presented. The purpose of the evaluation was to identify the effects on gauge measurement of various aggregate types and antistripping agents in the Marshall specimens. In addition, results of testing two Marshall specimens simultaneously were compared with those results obtained by testing specimens singly to determine which of the two is the preferred test method. To analyze the effects of aggregate-type and antistripping agent on the measurements, 675 four-minute gauge readings were obtained. The gauge measured asphalt content differently among groups of specimens composed of different aggregate types. None of the antistripping agents tested had a significant effect on the capacity of the gauge to consistently measure asphalt content. The results of the analysis show that consistent gauge readings are obtained if the gauge is recalibrated daily. The means and variances of 40 four-minute gauge readings were analyzed to determine if simultaneously testing two specimens yielded more accurate asphalt content measurements. The results of that analysis indicate that gauge readings of lower variances and higher accuracies are obtained by testing two specimens together. A recommended procedure for testing Marshall specimens with the asphalt content gauge is included. The procedure calls for recalibrating the gauge for each day of test, establishing different calibrations for each type of aggregate used in the specimens, and simultaneously testing two Marshall specimens.

Accurate control and measurement of asphalt content is one of the most fundamental requirements for producing high-quality bituminous mixtures. The determination of the asphalt content of a mixture by traditional methods is generally a time-consuming and somewhat hazardous procedure. The demand for a faster and safer method of asphalt content determination prompted the development of the nuclear asphalt content gauge. The nuclear gauge is used by state and federal engineering agencies, contractors, and materials laboratories for quality control and testing of bituminous mixtures.

J. L. Burati, Jr., Department of Civil Engineering, Clemson University, Clemson, S.C. 29634. T. E. Freeman, William F. Freeman Associates, Airpark West No. 2, Suite 100, 7823 National Service Road, Greensboro, N.C. 27409. H. S. Hill, Jr., Department of Experimental Statistics, Clemson University, Clemson, S.C. 29634.

The procedure for testing asphalt contents of bituminous mixtures with the nuclear gauge is standardized by the American Society for Testing and Materials (ASTM). The Troxler 3241 asphalt content gauge evaluated in this paper meets all requirements designated by ASTM D4125-83, *Standard Method of Test for Asphalt Content of Bituminous Mixtures By the Nuclear Method* (*1*).

Traditional extraction methods for measuring the asphalt content of a bituminous mixture involve (a) determination of the moisture content of the mixture, (b) physical separation of bitumen from aggregate, and (c) weighing each component to determine the percentage of bitumen in the test portion. ASTM D2172-31, *Standard Test Methods for Quantitative Extraction of Bitumen From Bituminous Paving Mixtures* (*2*), standardizes centrifuge, reflux, and vacuum extraction test procedures.

All of the extraction methods involve a succession of steps, most of which introduce risk of error. Consequently, the tests are very sensitive to operator technique and experience. None of the tests can be completed easily in less than 2 hr, and they can often take up to a full day to return results. In addition, each extraction test involves the use of very dangerous chemicals. Nonetheless, the accurate measurement of asphalt content in bituminous mixtures is absolutely necessary for quality control.

The benefits inherent in asphalt content determination of bituminous mixtures by the nuclear method are numerous. Time savings, reduced safety hazards, improved accuracy, and nondestructive testing are among these advantages. One problem, however, with testing asphalt content using a nuclear gauge is related to the preparation of bulky, non-standard samples of asphalt mix.

The Model 3241 gauge is designed to test a sample of hot asphalt mix of approximately 6,800 g. The large sample is particularly troublesome in the preparation of calibration samples for establishing the calibration curve for each asphalt mix. Some agencies currently using the gauge stress the importance of maintaining a constant weight for the calibration samples, while others stress the importance of a constant volume. The use of a smaller, standard-size sample that can achieve both the constant weight and volume requirements, while at the same time being capable of being tested for other asphalt properties as well as for asphalt content, would greatly contribute to the quality control and testing of bituminous mixtures. Specimens prepared in accordance with the Marshall method of mix design (*3*) meet all of the stated requirements.

The purpose of this research was to investigate the effectiveness of implementing the Marshall specimen as a standard test sample for asphalt content determination by the nuclear method using the Troxler 3241 asphalt content gauge. Specifically, the following objectives were addressed:

1. Analyze the effects on gauge performance of varying aggregate types and antistripping agents in the Marshall specimens.
2. Determine if the gauge can accurately measure asphalt contents of Marshall specimens.
3. Compare the observed asphalt content values based on daily gauge calibrations with those values obtained using an initial gauge calibration only.
4. Determine if gauge readings obtained by testing two Marshall specimens together vary less and are more accurate than those readings obtained by testing one specimen alone.

EXPERIMENTAL PROCEDURE, SINGLE SPECIMENS

The phase of research dealing with nuclear testing of single Marshall specimens was divided into two main tasks. The first task involved determining the accuracy with which the gauge measures the asphalt content of Marshall specimens. The first task also included an analysis to determine what effects various aggregate types and antistripping agents in the specimens have on gauge measurements. The second task involved comparing gauge readings based on new calibration curves for each day of testing with those readings obtained using a single calibration curve throughout the experiment.

Experimental Design

Four aggregate types, each with a different corresponding actual asphalt content, and four different antistripping agents, plus a control with no antistripping agent (a total of five agents) were used in preparing the test specimens. Other variables were held as constant as possible to prevent them from influencing gauge readings. Such variables included asphalt source and grade, temperature (mixing, compaction, and testing), equipment, and operators. All specimens composed of a particular aggregate type had the same asphalt content. A total of seven replicates were used. Each replicate contained 20 different specimens, 1 specimen for each possible aggregate type (four) and antistripping agent (five) combination, making a total of 135 (140 less 5 missing) Marshall specimens. Five gauge readings (observations) were obtained and averaged for each specimen to minimize error due to variation in gauge reading.

The randomized complete block design was selected as the experimental design because of its simplicity and its facility in minimizing experimental error due to variation between replicates. Because all replicates were not prepared in a single day (one replicate consisting of 20 specimens was prepared per day), the potential existed for variation between replicates. In this experiment, each replicate is a separate block, each Marshall specimen is an experimental unit, and each aggregate type and antistripping agent combination is a separate treatment. Hence, there are 7 blocks, 20 treatments, and 135 experimental units.

Specimen Preparation Procedure

Standard-size Marshall specimens, 2.5 in. tall and 4.0 in. in diameter, were prepared under controlled conditions in the materials laboratory at Clemson University.

Description of the Material Tested

The four types of aggregate used in preparing all calibration and test specimens were obtained from quarries located in South Carolina. These quarries were selected because the specimens were to be tested later in an asphalt stripping research project for the South Carolina Department of Highways and Public Transportation (SCDHPT), and that project required Marshall specimens composed of aggregate from the selected quarries. The aggregate gradations and asphalt cement quantities used to prepare the test specimens are based on SCDHPT specifications for asphaltic concrete overlays.

For this paper, the aggregate sources are designated as C, F, L, and P. Aggregate F was a coastal plains sand and gravel, while Aggregates C, L, and P were crushed granite, crushed granite-gneiss, and crushed granite, respectively. The asphalt cement was an AC-20 from a source commonly used in South Carolina. The test specimens contained the optimum asphalt content as determined by the SCDHPT mix design. The mixes for Aggregates C, F, L, and P contained 5.9, 5.6, 6.2, and 6.3 percent asphalt cement, respectively.

Specimens were prepared with no antistripping agents added, with three liquid antistripping agents widely used in South Carolina, and with hydrated lime added. The four antistripping agents (three liquid agents and lime) were randomly designated Agents 1, 2, 3, and 4, whereas the control specimens with no antistripping additives were designated Agent 0.

Mixing and Compaction Procedure

Mixing and compaction procedures were in accordance with Marshall procedures published by the Asphalt Institute (*3*) with the exception that a mechanical hammer was used and 20 compactive blows were applied to each face of the specimens. This number of blows produced specimens containing 6 to 8 percent air voids. This air void content satisfied the moisture susceptibility testing requirements for the SCDHPT stripping research mentioned earlier.

Gauge Testing Procedure

The procedure for testing was divided into two tasks. The first task involved establishing calibration curves by recording slopes and intercepts for each pair of calibration specimens. The second task involved recording measure counts and asphalt content determinations for the test specimens.

Calibration

A total of eight calibration specimens, two for each aggregate type, were prepared in accordance with the procedure used to fabricate the test specimens. No antistripping agent was used in the calibration specimens, and each specimen had a different asphalt content. The asphalt contents for the calibration specimens were determined according to the calibration procedure outlined in the *Model 3241 Instruction Manual* (*4*). For each aggregate type, one calibration specimen was prepared at 1 percent below design asphalt content and one was prepared at 1 percent above design asphalt content.

Four two-point calibration curves, one for each aggregate type, were established at the beginning of each day. This was accomplished by obtaining two 16-minute counts per aggregate type, one for the lower percent asphalt content specimen and one for the upper. The percentages of asphalt content for both specimens were entered into the gauge. At that time, the slope and intercept for the curve were computed, recorded, and stored in the gauge memory for access during testing.

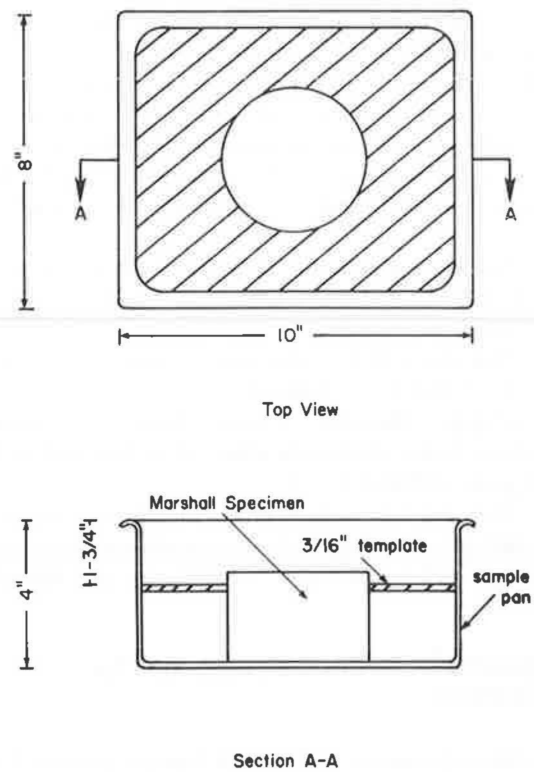

FIGURE 1 Sample pan and template used to align one Marshall specimen.

Testing

To ensure consistent alignment of the specimens in the gauge, a 3/16-in.-thick aluminum template was constructed. A 4-in.-diameter hole was cut in the template as shown in Figure 1. The template fit tightly into a standard sample pan, and the specimen was placed through the hole in the template and allowed to rest on the bottom of the pan.

Five gauge readings were obtained for all 20 specimens in a replicate before the next replicate was tested. The complete testing of one replicate, 100 gauge readings, required 2 days. Therefore, during a given day, either 40 or 60 gauge readings were obtained, thereby ensuring that the replicate would be wholly tested at the end of the day. The order of testing for a given day was established by randomly ordering the 20 specimens composing the replicate being tested. The random ordering was repeated before each round of tests. Before each specimen was tested, the calibration curve corresponding to that aggregate type was recalled from the gauge memory and established as the basis for computing its percent asphalt content. All gauge readings for test specimens were based on 4-minute counts. Test order, displayed measure counts, and asphalt content readings were recorded.

EXPERIMENTAL RESULTS: SINGLE SPECIMENS

Five test specimens were missing from the data set. Because each specimen was tested five times to obtain one average reading, 25 observations, in total, were missing. Therefore, 700 possible observations, minus the 25 missing ones, provided a total of 675 observations to be analyzed. The first step in analyzing the data was to average the five observations for each specimen. This process resulted in a total of 135 averaged readings.

Effect of Replicate, Aggregate Type, and Antistripping Agent on Gauge Performance

The percent asphalt content was subtracted from the gauge reading for each of the 135 specimens to determine the deviation of the observed from the actual asphalt content. These differences were sorted according to quarry and antistripping agent and an analysis of variance (ANOVA) using the general linear models (GLM) procedure of the Statistical Analysis System (SAS) (*5*) computer program was conducted to determine if replicate, quarry, or antistripping agent had a significant effect on the gauge readings.

Table 1 summarizes the results of the GLM procedure. The results show that the differences between replicates are not significant, therefore, groups of readings obtained in one time period do not differ from readings made at another time. The results also show that differences are not significant among antistripping agents, nor does the interaction of replicate by agent have a significant effect on the readings. Similarly, no significant effects result from the interactions of replicate by quarry or quarry by agent. Differences between quarries, however, are significant, as indicated by the low probability of obtaining an *F* value as large as the one computed.

TABLE 1 RESULTS OF ANOVA ON DIFFERENCES BETWEEN GAUGE READINGS AND ACTUAL ASPHALT CONTENTS CONSIDERING REPLICATES, QUARRIES, AND AGENTS (USING DAILY GAUGE CALIBRATIONS)

Source (or Factor)	Result +	Prob > F #
Replicate	NS	.512
Quarry	S	.001
Agent	NS	.302
Quarry*Agent ^	NS	.503
Replicate*Quarry ^	NS	.071
Replicate*Agent ^	NS	.360

\+ NS - Not significantly different at the 5% level
S - Significantly different at the 5% level

\# Probability of obtaining an F-value as large as the one computed if factor level means are actually equal.

^ Interaction terms between the variables shown

Analysis of Gauge Measurement Accuracy

The results presented in Table 1 indicate that the data can be regrouped to make further analysis possible. Specifically, the accuracy of the readings can be analyzed by comparing with a true mean of zero the mean difference between observed and actual asphalt content for the 20 quarry-agent combinations.

The differences between observed and actual percent asphalt content for the 135 readings were averaged over the seven replicates to obtain 20 mean differences, 1 for each quarry-agent combination. The UNIVARIATE procedure in SAS was used to analyze the 20 mean differences. As part of the UNIVARIATE procedure, a *t*-test is conducted to compare the mean differences to a true mean of zero to determine which, if any, of the mean differences are significantly different from zero.

Of the 20 quarry-agent combinations, only the C-0, C-1, and C-3 treatments had mean differences significantly different from zero. This indicates that the Marshall specimens can be tested with accurate results if the gauge is calibrated daily and if separate calibrations are established for different aggregate types.

It does not seem necessary, however, to recalibrate the gauge whenever a different antistripping agent is used in the specimens. That the mean difference between observed and actual percent asphalt content is positive for a vast majority of the treatments might suggest that the gauge has a tendency to detect slightly more asphalt in the specimens than is actually present. Table 2 presents the results of the UNIVARIATE analysis.

Analysis of Calibration Method

One objective of the research was to compare gauge readings that were obtained using four new calibration curves every day (one curve for each aggregate type) with readings based on the same four calibration curves throughout the test period. The results of this analysis can determine whether calibrating the gauge one time only for each aggregate type yields sufficiently consistent gauge readings for Marshall specimen testing. The calibration procedure required approximately 3 hours per day.

Because measure counts and pertinent calibration data, as well as gauge readings, were recorded, it was relatively simple to recompute the 675 equivalent percent asphalt content readings that would have been displayed by the gauge had the original four calibration curves been used to compute the readings. This recomputation was accomplished by using the slope-intercept equation for a straight line and solving for y:

$$y = mx + b$$

where

y = percent asphalt content reading based on original calibration,
m = slope of original calibration curve,
x = measure count recorded during the test, and
b = vertical intercept of the original calibration curve.

The actual asphalt contents were then subtracted from each of the 675 calculated readings to determine the deviations of the observed readings from the actual contents. The 5 calculated differences for each of the test specimens were averaged to obtain 135 averaged differences. These differences were sorted according to quarry and antistripping agent, and an ANOVA was conducted to determine if the effect of replicates had any statistical significance in the differences.

TABLE 2 RESULTS OF *t*-TESTS COMPARING MEAN DIFFERENCES BETWEEN GAUGE READINGS AND ACTUAL ASPHALT CONTENT TO A TRUE MEAN OF ZERO (USING DAILY GAUGE CALIBRATIONS)

Treatment	Actual % Asphalt	Mean Gauge Reading	Mean Reading-Actual % A.C.	Result*	Prob > |t|#
Agg-Agent					
C-0	5.9	6.01	.11	S	.038
C-1	5.9	6.05	.15	S	.004
C-2	5.9	5.94	.04	NS	.607
C-3	5.9	6.02	.12	S	.032
C-4	5.9	5.96	.06	NS	.359
F-0	5.6	5.67	.07	NS	.245
F-1	5.6	5.69	.09	NS	.126
F-2	5.6	5.74	.14	NS	.077
F-3	5.6	5.67	.07	NS	.481
F-4	5.6	5.61	.01	NS	.811
L-0	6.2	6.23	.03	NS	.292
L-1	6.2	6.25	.05	NS	.245
L-2	6.2	6.23	.03	NS	.600
L-3	6.2	6.19	-.01	NS	.785
L-4	6.2	6.18	-.02	NS	.708
P-0	6.3	6.20	-.10	NS	.169
P-1	6.3	6.30	.00	NS	.910
P-2	6.3	6.38	.08	NS	.188
P-3	6.3	6.28	-.02	NS	.716
P-4	6.3	6.30	.00	NS	.947

* NS - Not significantly different at the 5% level
S - Significantly different at the 5% level

Probability of obtaining a t-value as large as the one computed if the mean difference is actually zero.

Results in Table 3 show that the probability of obtaining an F value as large as the one computed is very small for the replicate source, indicating that the differences between replicates are significant when only an initial calibration curve is used to measure the readings. Because replicates represent different blocks of time, it appears that sufficiently consistent gauge readings to test Marshall specimens are not obtained when the same calibration curve is used for extended periods of time. In this experiment, all seven replicates were tested over a period of 18 days.

EXPERIMENTAL PROCEDURE: DOUBLE SPECIMENS

It was hypothesized that the simultaneous testing of two specimens would produce less variable gauge measurements than testing single specimens alone. The Model 3241 gauge was designed to test asphalt samples of substantially large masses. Two Marshall specimens together more closely approximate the large samples of asphalt mix that the gauge was designed to test. Some of the variation between readings that occurs when testing a single specimen may be due to the lack of sufficient sample mass for the gauge counters to detect slowed hydrogen neutrons. The purpose of this phase of research was to determine whether testing two specimens together produces more consistent results than testing a single specimen alone.

Experimental Design

The same four aggregate types used in the first phase of the research were also used in the second phase. Asphalt source and grade, temperatures (of mixing, compaction, and testing), equipment, and operators were held as constant as possible to prevent them from influencing the gauge readings. This phase of the research used a total of four replicates. Each replicate contained two test specimens, making a total of eight specimens to be tested.

When two specimens are tested together, each is positioned above one of the two detectors contained in the gauge. The possibility that the detectors count slowed hydrogen neutrons differently presented an interesting problem for this phase of research, namely, that the position of the two specimens in the gauge may have a significant effect on the capacity of the

TABLE 3 RESULTS OF ANOVA ON DIFFERENCES BETWEEN GAUGE READINGS BASED ON INITIAL CALIBRATIONS AND ACTUAL ASPHALT CONTENTS CONSIDERING REPLICATES, QUARRIES, AND AGENTS

Source (or Factor)	Result +	Prob > F #
Replicate	S	.0001
Quarry	S	.003
Agent	NS	.245
Quarry*Agent ^	NS	.298
Replicate*Quarry ^	S	.0001
Replicate*Agent ^	NS	.262

+ NS - Not significantly different at the 5% level
  S - Significantly different at the 5% level

Probability of obtaining an F-value as large as the one computed if factor level means are actually equal.

^ Interaction terms between the variables shown

gauge to consistently detect slowed neutrons. To eliminate the variation in readings due to possible dissimilar neutron detection, the specimens were tested an equal number of times over each detector. The experiment was designed so that an analysis of the effect on gauge readings of the relative specimen positions could be made.

The randomized complete block was selected as the design for this experiment. Each aggregate type is a separate block, or replicate, and four treatments were applied to each block. Ten readings, also known as subsamples, were obtained for each treatment. Therefore, 40 observations were made for each block, yielding a total of 160 gauge readings to be analyzed for the experiment. For a given block, the four possible arrangements, or positions in the gauge of the two test specimens A and B, constitute the treatments. Treatment 1 is defined as Specimen A tested alone. Treatment 2 is defined as Specimen B tested alone. Treatment 3 is defined as Specimen A in the left side of the sample pan above one detector with Specimen B in the right side of the sample pan above the other detector. Treatment 4 is the arrangement of Specimen B in the left and Specimen A in the right. The experimental design is presented in Table 4.

TABLE 4 EXPERIMENTAL DESIGN FOR SINGLE- VERSUS DOUBLE-SPECIMEN TESTING

	Block 1	Block 2	Block 3	Block 4
QUARRY	C	F	L	P
Treatment				
1	A	A	A	A
2	B	B	B	B
3	AB	AB	AB	AB
4	BA	BA	BA	BA

LEGEND: A - Specimen A tested alone
B - Specimen B tested alone
AB - Specimens A and B tested together with A in left, B in right
BA - Specimens A and B tested together with B in left, A in right

F-tests and t-tests can be used to compare variances and means between single- and double-specimen readings. The mean squared error, defined as standard deviation squared plus bias squared, can be computed for groups of single- and double-specimen readings. The test method that yields the lower mean squared error is the preferred method.

Specimen Preparation Procedure

Four calibration specimens and two test specimens were prepared for each aggregate type. All calibration and test specimens were prepared in a single day in accordance with the same procedure used to prepare the specimens analyzed in the first phase of the research with one exception. No antistripping agents were applied in the preparation of these specimens.

Gauge Testing Procedure

Calibration

It was necessary to calibrate the gauge with two specimens below (low) and two above (high) design asphalt content for each aggregate type in order to be consistent with the test procedure. Therefore, for the same reasons previously discussed, the relative specimen positions had to be considered during calibration. One slope and intercept was obtained for each of the four possible double-calibration specimen arrangements and averaged together to yield one calibration curve that would be used to test all double specimens of a given aggregate. Similarly, the single-specimen calibration curve used for testing was the average of four individual slopes and intercepts obtained from the four possible combinations of low-high specimen arrangements. Table 5 more clearly presents the calibration procedure.

Testing

Because of complications in the calibration procedure, only one aggregate type was tested each day. Ten observations per treatment were obtained each day, making a total of forty 4-min gauge readings. Eight calibrations based on 16-min counts were made and averaged each day to obtain two useful calibration curves. The order of testing for each day was determined randomly. Before testing, the appropriate calibration curve (single or double) was recalled from the gauge memory and established as the basis for computing asphalt content for that test. The two specimens were held in position with an aluminum template containing two 4-in.-diameter holes. The template fit tightly in the sample pan (see Figure 2). The measure counts and displayed asphalt contents were recorded for each test.

EXPERIMENTAL RESULTS: DOUBLE SPECIMENS

The first step in analyzing the data was to sort the observations according to quarry and treatment. The actual asphalt content was subtracted from each observation to determine the deviation of the gauge readings from the actual content.

TABLE 5 SINGLE- VERSUS DOUBLE-SPECIMEN CALIBRATION PROCEDURE FOR A GIVEN AGGREGATE

Calibration Curve	Pan #1 Contents	Pan #2 Contents	End Product
	Single Specimen Calibration		
1	A-Low	A-High	Slope 1 Intercept 1
2	A-Low	B-High	Slope 2 Intercept 2
3	B-Low	A-High	Slope 3 Intercept 3
4	B-Low	B-High	Slope 4 Intercept 4
	Double Specimen Calibration		
5	A-Low B-Low	A-High B-High	Slope 5 Intercept 5
6	A-Low B-Low	B-High A-High	Slope 6 Intercept 6
7	B-Low A-Low	A-High B-High	Slope 7 Intercept 7
8	B-Low A-Low	B-High A-High	Slope 8 Intercept 8

NOTE: One curve used for single and one curve used for double specimen testing are obtained by averaging slopes and intercepts corresponding to calibration curves 1-4 and curves 5-8, respectively.

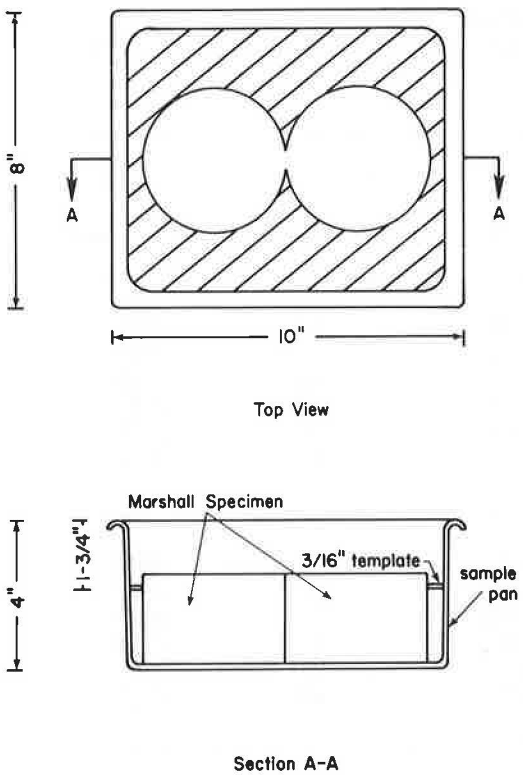

FIGURE 2 Sample pan and template used to align two Marshall specimens.

Analysis of Gauge Measurement Accuracy

The UNIVARIATE procedure generated t-tests to determine whether mean differences between observed and actual asphalt contents are significantly different from zero. The comparison to a true mean of zero was made for each of the treatments individually and for Treatments 1 and 2 combined (single specimens), as well as for Treatments 3 and 4 combined (double specimens).

Results of the analysis presented in Table 6 indicate that the gauge readings for each treatment are close to the actual asphalt contents. The combined readings of Treatments 1 and 2 and Treatments 3 and 4 are accurate, as indicated by the relatively high probabilities of obtaining t values as large as the ones computed. Only the combined readings of Treatments 3 and 4 for Aggregate P have mean differences that are significantly different from zero at the 5 percent level.

Comparison of Readings Between Treatments

It is of value to know if Treatment 1 readings are significantly different from Treatment 2 readings, and likewise, if Treatment 3 readings are different from those of Treatment 4, for a common aggregate type. If the gauge readings are not significantly different between treatments, it is possible to group readings from the two treatments, thus enabling analysis of a larger sample size. Student's t-tests were conducted to make comparisons between the 10 Treatment 1 readings and the 10 Treatment 2 readings for each aggregate type. The t-tests were similarly conducted on the readings from Treatments 3 and 4. F-tests were conducted on the same data to determine if variances between treatments were significantly different. The null hypothesis tested in all cases was that the two treatment means (or variances) were equal.

The results indicate that there is no significant difference at the 5 percent level between Treatments 1 and 2 mean readings, nor is there a significant difference between the mean readings of Treatments 3 and 4 for any of the aggregate types. These results make it possible to group Treatment 1 readings with Treatment 2 readings and Treatment 3 readings with Treatment 4 readings, thereby enabling further analysis of larger sample sizes. The lack of significant difference between Treatment 3 readings and Treatment 4 readings indicates that the position of the two test specimens in the gauge has no effect on the capacity of the gauge to consistently detect slowed hydrogen neutrons. It is now possible to test for equality of variances among all 20 single-specimen and all 20 double-specimen readings for each aggregate type (see Table 7).

Comparison of Means and Variances Between Single- and Double-Specimen Readings

To determine if the means and, more important, the variances between single- and double-specimen readings are significantly different, t-tests and F-tests comparing the means and variances of the 40 observations in each block were conducted. The bias of the readings, defined as the deviation of the observed from the actual asphalt content, is not, by itself, sufficient information to determine which of the test methods is better. Nor is variance alone a good indicator of the better test method. The sum of variance and the square of the bias, known as the mean squared error, can give a good indication of the better test method.

Table 8 presents the results of this analysis. None of the mean readings is significantly different between the two test methods for any of the aggregate types. The variances, however, are significantly smaller for all aggregate types when the double-specimen test method is used. This difference verifies the hypothesis that the gauge readings vary less when two specimens are tested together than when one is tested alone. The bias is smaller for double-specimen readings in all cases except for those readings obtained using specimens of Aggregate P. The mean squared error term is smaller for readings obtained using the double-specimen test method for all aggregate types. The smaller mean squared error terms demonstrate that the double-specimen test method produced gauge readings that were less variable and more accurate than those readings produced by testing a single Marshall specimen.

CONCLUSIONS AND RECOMMENDATIONS

The following conclusions were reached from the laboratory results concerning the first phase of research, the nuclear

TABLE 6 RESULTS OF t-TESTS COMPARING MEAN DIFFERENCES BETWEEN GAUGE READINGS AND ACTUAL ASPHALT CONTENT TO A TRUE MEAN OF ZERO FOR TREATMENTS 1 AND 2 (SINGLE SPECIMENS) AND TREATMENTS 3 AND 4 (DOUBLE SPECIMENS)

Treatment	Actual % Asphalt	Mean Gauge Reading	Mean Reading-Actual % A.C.	Result*	Prob > |t|#
Aggregate C					
1	5.9	5.85	-.05	NS	.538
2	5.9	5.80	-.10	NS	.209
3	5.9	5.90	.00	NS	.926
4	5.9	6.00	.10	NS	.115
1&2^	5.9	5.82	-.08	NS	.172
3&4^	5.9	5.95	.05	NS	.171
Aggregate F					
1	5.6	5.63	.03	NS	.607
2	5.6	5.66	.06	NS	.467
3	5.6	5.62	.02	NS	.624
4	5.6	5.66	.06	NS	.124
1&2^	5.6	5.65	.05	NS	.354
3&4^	5.6	5.64	.04	NS	.131
Aggregate L					
1	6.2	6.18	-.02	NS	.626
2	6.2	6.11	-.09	NS	.408
3	6.2	6.21	.01	NS	.736
4	6.2	6.21	.01	NS	.773
1&2^	6.2	6.14	-.06	NS	.324
3&4^	6.2	6.21	.01	NS	.660
Aggregate P					
1	6.3	6.37	.07	NS	.181
2	6.3	6.22	-.08	NS	.371
3	6.3	6.33	.03	NS	.309
4	6.3	6.36	.06	NS	.102
1&2^	6.3	6.30	.00	NS	.937
3&4^	6.3	6.34	.04	S	.049

* NS - Not significantly different at the 5% level
S - Significantly different at the 5% level

\# Probability of obtaining a t-value as large as the one computed if the mean difference is actually zero.

^ 1&2 - Treatments 1 and 2 combined
3&4 - Treatments 3 and 4 combined

testing of Marshall specimens, and the second phase of research, single- versus double-specimen testing.

1. It is possible to determine asphalt contents of Marshall specimens using the Model 3241 nuclear gauge and obtain results that are not significantly different from the specimens' actual asphalt contents at the 5 percent level of significance, provided the gauge is recalibrated daily and separate calibration curves are provided for different aggregate types.
2. The presence of antistripping agents used in this study in the specimens had no significant effect on the capacity of the gauge to measure the asphalt content of Marshall specimens.
3. The gauge may measure asphalt content differently between specimens composed of different aggregate types.
4. Gauge readings are significantly different between replicates of specimens if the same calibration curve is used to test a series of replicates.
5. Gauge readings with smaller mean squared error terms were obtained by testing two Marshall specimens simultaneously rather than testing one specimen alone. This indicated that readings with generally larger accuracies and

TABLE 7 HYPOTHESIS TEST RESULTS COMPARING MEANS AND VARIANCES BETWEEN TREATMENTS 1 AND 2 GAUGE READINGS (SINGLE SPECIMENS) AND TREATMENTS 3 AND 4 GAUGE READINGS (DOUBLE SPECIMENS)

Treatment	Actual % Asphalt	Mean Reading	Result* (Prob >|t|)#	Variance	Result* (Prob > F)^
Aggregate C					
1	5.9	5.85	NS	.071	NS
2		5.80	(.661)	.060	(.809)
3	5.9	5.90	NS	.010	NS
4		6.00	(.136)	.032	(.098)
Aggregate F					
1	5.6	5.63	NS	.036	NS
2		5.66	(.787)	.060	(.454)
3	5.6	5.62	NS	.011	NS
4		5.66	(.383)	.013	(.834)
Aggregate L					
1	6.2	6.18	NS	.021	NS
2		6.01	(.109)	.081	(.056)
3	6.2	6.21	NS	.008	NS
4		6.21	(.943)	.022	(.159)
Aggregate P					
1	6.3	6.37	NS	.023	NS
2		6.22	(.140)	.069	(.123)
3	6.3	6.33	NS	.009	NS
4		6.36	(.567)	.010	(.872)

* NS - Not significantly different at the 5% level
S - significantly different at the 5% level

\# Probability of obtaining a t-value as large as the one computed if the means are actually equal

^ Probability of obtaining an F-value as large as the one computed if the variances are actually equal

smaller variances were obtained using the double-specimen test method.

6. The configuration of the two specimens in the gauge had no significant effect on the capacity of the gauge to consistently measure asphalt content.

These conclusions are applicable for the conditions, that is, the asphalt cement, asphalt contents, aggregate types and gradations, and test time used in this laboratory investigation.

Based on the research findings, the following laboratory procedure is suggested for asphalt content determination of Marshall specimens using the asphalt content gauge.

Prepare two calibration specimens with asphalt contents approximately 1 percent below design asphalt content of the mixture being tested. Prepare two specimens with asphalt contents approximately 1 percent above the design asphalt content. Using an aluminum template to ensure consistent alignment of the specimens in the sample pan, calibrate the gauge by making two 16-minute measure counts. If possible, calibrate the gauge just prior to testing the mixture in question. Establish a separate calibration curve for each mixture composed of a different aggregate type.

Prepare two Marshall specimens from the mixture to be tested. After the specimens have cooled to room temperature, place them in the gauge, noting their positions, and obtain one 4-minute reading. Reverse the positions of the specimens and make another 4-minute test. Average the two observations to obtain one measured asphalt content.

TABLE 8 MEAN SQUARED ERROR CALCULATIONS AND HYPOTHESIS TEST RESULTS COMPARING MEANS AND VARIANCES BETWEEN ALL SINGLE- AND DOUBLE-SPECIMEN GAUGE READINGS

1 Specimen Test Method	2 Actual % Asphalt	3 Mean Reading	4 Variance	5 Bias *	6 Mean Squared Error *
Aggregate C					
Single	5.9	5.82	.063	-.08	.069
Double	5.9	5.95	.022	.05	.024
RESULT #		NS (.061) ^	S (.029) +		
Aggregate F					
Single	5.6	5.65	.046	.05	.050
Double	5.6	5.65	.011	.05	.013
RESULT #		NS (.882) ^	S (.004) +		
Aggregate L					
Single	6.2	6.14	.065	-.06	.069
Double	6.2	6.21	.014	.01	.014
RESULT #		NS (.278) ^	S (.002) +		
Aggregate P					
Single	6.3	6.30	.049	.00	.049
Double	6.3	6.35	.010	.05	.012
RESULT #		NS (.365) ^	S (.001) +		

* Col 5 = Col 3 - Col 2
Col 6 = Col 4 + $(\text{Col } 5)^2$

\# RESULT - Hypothesis test result:
NS - Not significantly different at the 5% level
S - Significantly different at the 5% level

^ Probability of obtaining a t-value as large as the one computed if the means are actually equal

\+ Probability of obtaining an F-value as large as the one computed if the variances are actually equal

ACKNOWLEDGMENTS

Acknowledgment is given to Troxler Electronic Laboratories, Inc., for its sponsorship of this project through the Department of Civil Engineering of Clemson University. Special thanks are given to Glen Watters and Yon Larrea of Troxler Electronic Laboratories, Inc., for their assistance and continuous support throughout the investigation.

REFERENCES

1. *Standard Method of Test for Asphalt Content of Bituminous Mixtures By the Nuclear Method.* ASTM D4125-83. Annual Book of ASTM Standards, Section 4, ASTM, Philadelphia, Pa., 1984, p. 688.
2. *Standard Test Methods for Quantitative Extraction of Bitumen From Bituminous Paving Mixtures.* ASTM D2172-81. Annual Book of ASTM Standards, Section 4, ASTM, Philadelphia, Pa., 1984, pp. 350-364.
3. *Mix Design Methods for Asphalt Concrete and Other Hot-Mix Types.* The Asphalt Institute Manual Series MS-2. College Park, Md., March 1979.
4. *Model 3241 Instruction Manual.* Troxler Electronics Laboratories, Inc., Research Triangle Park, N.C., 1983, p. 1.
5. *SAS User's Guide: Basics, 1982 Edition.* SAS Institute Inc., Cary, N.C., 1982.

Publication of this paper sponsored by Committee on Characteristics of Bituminous Paving Mixtures to Meet Structural Requirements.