Comparison of End Result and Method Specifications for Managing Quality

PAUL E. BENSON

The results of a statistically designed experiment in which asphalt concrete cores and nuclear gauge readings were taken from five California projects are reported. Relative compaction for the projects was controlled with a method specification. Analysis of variance is used to separate test error and locational components of variance for specific gravity, asphalt content, air voids, lift thickness, and grading. Compaction results are compared with similar results from 16 end result specification (ERS) jobs studied previously. Relative compaction on the ERS jobs averages 3.1 percent higher in value. Findings on test precision, increased lot size, and materials variability are discussed.

Prescriptive, or method, specifications have been used for many years by transportation agencies to control quality. Such specifications typically are applied to materials for which significant lapses in quality would require removal and replacement. Compaction of asphalt concrete is one of the most widespread examples of this type of application.

In theory, method specifications reduce job delays, contract claims, and escalation in future bid prices by ensuring that the work is done right the first time. However, the prescriptive approach has two important disadvantages. First, it stifles innovation and competitiveness by prescribing exactly how the work is to be done. Second, it requires the full-time presence of experienced field personnel for proper enforcement.

In response to these problems, many states have elected to implement end result specifications (ERS) combined with pay adjustment factors. These measures make contractors responsible for achieving quality but let them decide how to do it. Doing this makes sense from economic as well as contractual standpoints. Contractors are in a better position to manage the day-to-day quality of their product because of their direct involvement with suppliers and subcontractors and their direct control over construction activities. They are better able to experiment with new construction methods and will do so if it offers the possibility of a competitive advantage. The overall result is, theoretically, a high-quality product that meets design expectations.

In this paper, the validity of this theory is tested. To accomplish this, relative compaction data for asphalt concrete on jobs run under both specification systems are compared. As a by-product of this work, information on materials and testing variability for compaction and related mix properties is also presented.

EXPERIMENT DESIGN

California Department of Transportation (Caltrans) project engineers can opt for either a method or an end result specification to control the compaction of asphalt concrete. The method specification is still the standard control strategy, but a standard special provision for end result control is also available (*1*). This specification involves the use of nuclear gauge measurements and pay adjustment factors.

Since jobs using both specifications are being advertised and constructed in California, there is an opportunity to compare the level of quality achieved under each. In some cases, the performance for the same contractor or supplier can be compared. Since measurements of relative compaction are not available for the method specification jobs, a special effort was needed to obtain them. Five projects were chosen for coring and nuclear gauge measurements; their results are compared with data from 16 ERS jobs studied previously (*2*).

The method specification jobs were selected on a statewide basis to represent a range of job sizes and to include, to the extent possible, contractors who had also worked on ERS jobs. Jobs in Districts 3, 4, 5, 6, and 11 were chosen (Figure 1). The measurement program included cores as well as gage readings. The cores were needed to establish the test maximum density, but they were also useful in verifying the accuracy of the gage readings. Additional cores were taken for hot solvent extraction so that asphalt content and grading could be studied.

The experiment design was constructed as a full-factorial analysis of variance (ANOVA) with a fixed main effect (transverse location), two randomized main effects (gauge and longitudinal location), and a replication level of two. The model equation for nuclear gauge measurements X_{ijkl} is

$$X_{ijkl} = u - T_i + L_j + G_k + TL_{ij} + e_{l(ijk)} \quad (1)$$

where

u = true population mean,
T_i = transverse location (i = 1 to 3),
L_j = longitudinal location (j = 1 to 6),
G_k = nuclear gage (k = 1 to 2),
TL_{ij} = transverse-longitudinal interaction, and
$e_{l(ijk)}$ = experimental error term.

The only interaction term that was consistently significant in the analysis was TL_{ij}. The three gage interaction terms were pooled into the error term. For other test methods, the G_k term drops out of the equation.

Each job was divided into three sublots of equal length (Figure 2). Two longitudinal locations were selected randomly within each sublot for a total of six. The fixed transverse locations at each longitudinal station were at 0.3, 1.8, and 2.7 m (1, 6, and 9 ft) from the edge that was unsupported during the paving operation. Four 10-cm (4-in.) cores were taken at each test site on 25-cm (10-in.) centers

California Department of Transportation, 5900 Folsom Boulevard, Sacramento, Calif. 95819.

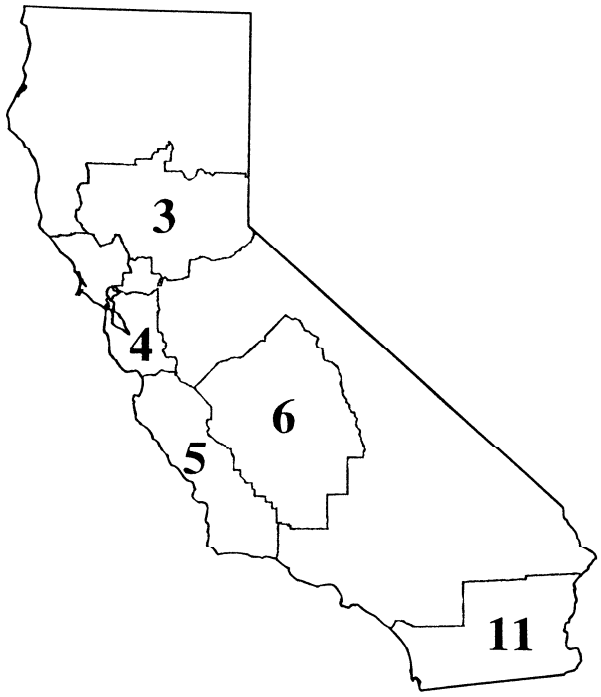

FIGURE 1 Participating districts.

in the direction of the paving operation. The first two, called A cores, were used to determine the in-place core and test maximum densities at each site. The second two, or B cores, were used in the hot solvent extraction portion of the study.

Within and between gauge components of variance were determined by using two nuclear gauges on each project. The first gauge, called the lab gauge, was used throughout the study. The second gage, called the district gauge, was supplied by the host district. Each gage had its own operator except the District 4 project, on which a single operator ran both gauges. Gage readings were taken at the two A core locations. All gauge measurements were made in the backscatter mode in accordance with California Test Method (CTM) 375.

The only other exception to the overall design was for the first job studied, District 11. On this job, no sublots were assigned. Both longitudinal and transverse locations were randomized. Five longitudinal stations and two transverse locations were selected. The model equation is the same for this design, but the expected mean square (EMS) table for the ANOVA is somewhat different.

All the method specification jobs were either on new alignment or in stage construction so that they were not trafficked before testing. Each job drew material from a single source except District 5, for which two sources were used. The study limits were defined as the final lift placed in a relatively continuous operation extending, at most, over 2 weeks. In contrast to this, data from the ERS jobs included measurements for all lots of asphalt concrete placed during an entire contract. If significant, this difference would tend to favor only higher relative compactions on the method specification jobs because of the additional attention that inspectors typically pay to placement of the final lift.

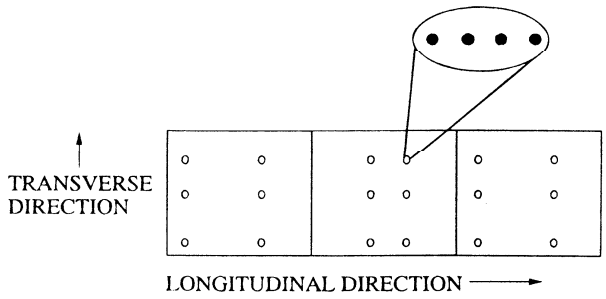

FIGURE 2 Sampling plan schematic.

LABORATORY TESTS

All laboratory tests on the field cores were performed at the headquarters laboratory in Sacramento by a crew of experienced technicians. Tests were completed within 9 months of the date of coring. Cores were stored in a controlled environment until testing. The cores were identified to the technicians by project number, test series (i.e., A or B), and core number (a random number assigned in the field). Neither the technicians nor their supervisors knew the location coordinates of the cores or the identity of the replicates. Testing proceeded in the order of the randomly assigned core numbers so that the effect of the test sequence was randomized over all factors.

The test methods performed in the laboratory are as follows:

- A Cores
  - –CTM 308: Methods of Test for Bulk Specific Gravity and Weight per Cubic Foot of Bituminous Mixtures
  - –ASTM D2041: Standard Test Method for Theoretical Maximum Specific Gravity and Density of Bituminous Paving Mixtures
- B Cores
  - –CTM 310: Method of Test for Determination of Asphalt and Moisture Contents of Bituminous Mixtures by Hot Solvent Extraction
  - –CTM 202: Sieve Analysis of Fine and Coarse Aggregates

Core densities were determined using the water-displacement method, which is Method C in CTM 308. Two operators worked as a team performing the density tests on the A cores, and one operator performed all the B core tests. The hot solvent extractions were run using four extractors. This equipment effect was randomized and is confounded in the test error reported for the percentage asphalt determination.

In preparation for testing, lift thickness was determined visually and measured for the A cores. The top 4.5 cm (0.15 ft), or less if the lift thickness fell below this value, was sawed off and used for the rest of the test sequence. The cores were dried to a constant weight at 38°C (100°F), usually taking 4 to 5 days, before specific gravity measurements were made. They were then heated to breakdown at 110°C (230°F), recompacted, allowed to cool, and tested again for specific gravity. Last, the maximum theoretical density was determined using ASTM D2041. The B cores were prepared for testing in the same manner as the A cores. After extractions were complete, a grading analysis was performed on the residual aggregate.

DATA ANALYSIS

All the pertinent field and laboratory measurements from the five method specification jobs were entered carefully into a computerized data base. The randomized site numbers were used to unscramble the data so that outliers could be examined using the difference between paired replicates as an indicator. This procedure turned up a small percentage of outliers, that were then examined for errors in transcription, calculation, or any other plausible explanations. Where mistakes were found, the appropriate corrections were made and entered into the data base. If no explanation could be found, the data were retained. In a few cases, insufficient material had been available to conduct a valid test. These results were removed from the data base and treated as missing data with a corresponding reduction in degrees of freedom for the ANOVA. Of the nearly 3,000 measurements made, only 20 were removed from the data base. At most, four measurements out of a possible total of 164 were removed for any given test method.

Results from the original District 11 gage had to be withdrawn from the analysis because the gage would not seat properly on the pavement. The error variance for this gage was nearly two orders of magnitude higher than any other gage used in the study. Fortunately, a third, experimental thin-lift gage had been used on this job as part of another study. Results from it were substituted for the district gage.

A summary of the data is given in Table 1, showing the averages by job for each of the tests, the overall average for all jobs, and the coefficients of variation attributable to both testing and materials.

Variance Components

To explore the significance of the experimental factors and partition the components of variance attributable to each, an ANOVA was run for each test method and project. The components of variance were computed from EMS tables based on the model equation and the types and levels of each factor involved (*3*). For the gage results from Districts 3 through 6, the components were determined from the following equations:

$$v_e = MS_e \tag{2}$$

$$v_G = (MS_G - MS_e)/36 \tag{3}$$

$$v_T = (MS_T - MS_{TL})/12 \tag{4}$$

$$v_L = (MS_L - MS_e)/6 \tag{5}$$

$$v_{TL} = (MS_{TL} - MS_e)/6 \tag{6}$$

where v is the component of variance and MS is the mean square (sum of squares/degrees of freedom) for the subscripted factor. Similar computations were made for all other data sets.

The components of variance from the individual projects were pooled into a single value for each test method. A summary of these pooled results is given in Table 2. Individual components of variance can be computed by taking the appropriate percentage of the total variance.

TABLE 1 Data Summary: Job Averages and Coefficients of Variation

	Average by District						Coefficient of Variation (%)	
Measurement Descriptions	3	4	5	6	11	Average All Jobs	Testing	Materials
Specific Gravity								
Original Core (Method C)	2.327	2.229	2.138	2.233	2.214	2.228	0.54	2.24
Nuclear Gage	2.229	2.116	2.081	2.190	2.150	2.153	1.89	3.51
Recompacted Cores (Method C)	2.429	2.393	2.217	2.349	2.342	2.346	0.67	1.22
ASTM D-2041	2.553	2.539	2.368	2.471	2.398	2.466	0.44	0.72
Relative Compaction (%)								
Cores (Method C)	95.81	93.02	96.42	95.07	94.57	94.98	0.61	1.95
Nuclear Gage	91.77	88.26	93.87	93.23	91.85	91.80	1.79	3.43
Original Cores								
Percent Air Voids (ASTM D-2041)	8.83	12.20	9.76	9.61	7.65	9.61	6.33	19.21
Asphalt Content (%)	5.18	4.76	5.55	5.19	6.14	5.36	3.50	4.36
Lift Thickness (cm)	6.11	5.27	5.01	4.52	5.80	5.34	4.27	17.87
Grading (% Passing)								
3/4" Sieve	99.7	99.3	98.8	98.7	100.0	99.3	1.2	0.7
1/2" Sieve	86.7	85.8	86.4	85.0	98.2	88.4	2.5	2.9
3/8" Sieve	77.3	70.5	73.1	72.1	89.2	76.4	2.8	5.1
#4 Sieve	60.3	53.3	49.6	54.8	62.4	56.1	2.6	7.2
#8 Sieve	43.9	38.3	38.0	43.4	47.5	42.2	2.6	8.6
#16 Sieve	34.6	26.4	30.6	35.6	36.1	32.7	2.3	9.4
#30 Sieve	25.6	16.3	21.9	23.0	26.1	22.6	2.4	9.5
#50 Sieve	16.5	9.7	12.2	11.5	16.9	13.4	2.7	8.8
#100 Sieve	10.7	6.0	6.6	5.6	10.1	7.8	3.6	9.9
#200 Sieve	7.4	4.3	4.3	3.5	6.6	5.6	4.8	11.1

TABLE 2 Components of Variance Summary: Pooled Results

Measurement Descriptions	Overall Std. Dev.	Total Variance	Source of Variation By Percent			
			Test Error	Transverse Location	Longitudinal Location	Transverse/ Longitudinal Interaction
Specific Gravity						
Original Core (Method C)	5.14E-02	2.64E-03	5.5	4.7	50.3	39.5
Nuclear Gage	8.59E-02	7.38E-03	22.4	13.3	45.8	18.5
Recompacted Cores (Method C)	3.27E-02	1.07E-03	23.4	1.0	68.4	7.2
ASTM D-2041	2.09E-02	4.36E-04	26.7	0.0	46.3	27.0
Relative Compaction (%)						
Cores (Method C)	1.94E+00	3.76E+00	8.9	14.7	35.7	40.7
Nuclear Gage	3.55E+00	1.26E+01	21.4	19.1	41.0	18.5
Original Cores						
Percent Air Voids (ASTM D-2041)	1.94E+00	3.78E+00	9.8	8.3	44.4	37.5
Asphalt Content (%)	3.00E-01	9.00E-02	39.2	4.6	41.3	14.9
Lift Thickness (cm)	9.81E-01	9.63E-01	5.4	6.8	60.8	27.0
Grading (% Passing)						
3/4" Sieve	1.39E+00	1.94E+00	74.5	8.9	7.8	8.8
1/2" Sieve	3.38E+00	1.14E+01	43.0	0.0	35.0	22.0
3/8" Sieve	4.40E+00	1.94E+01	23.2	2.9	51.5	22.4
#4 Sieve	4.31E+00	1.86E+01	11.6	3.2	65.4	19.8
#8 Sieve	3.79E+00	1.44E+01	8.1	3.2	75.7	13.0
#16 Sieve	3.16E+00	1.00E+01	5.6	2.4	81.1	10.9
#30 Sieve	2.22E+00	4.93E+00	5.8	1.5	81.4	11.3
#50 Sieve	1.24E+00	1.53E+00	8.4	1.4	78.6	11.6
#100 Sieve	8.23E-01	6.78E-01	11.5	2.0	73.4	12.1
#200 Sieve	6.79E-01	4.61E-01	16.0	1.0	70.2	12.8

The coefficients of variation presented in Table 1 are based on v_e as the testing component and the sum of v_T, v_L, and v_{TL} as the materials component. For the nuclear gage, v_e and v_c are summed to compute the testing component. Approximately one-third of this test error was due to repositioning and within-gage error, the rest being caused by differences between gages. For all other test methods, the effect of multiple test devices, if any, was confounded in v_e.

Because the replicate cores, although taken adjacent to each other, were not identical samples, there was concern that v_e might be an overestimate of test error. A Bartlett's test between the project error variances for original core densities rejected, at the 95 percent confidence level, the null hypothesis that the variances came from the same population (*4*). This meant that v_e was not consistent from project to project. However, the Spearman rank correlation coefficient between v_e and v_L for the five projects was insignificant, meaning that the difference in density introduced by the 25-cm (10-in.) offset between cores did not correlate with the macrolevel measure of longitudinal variability (*5*). It was concluded that the additional test error introduced by the coring offset was random and unavoidable given the nature of the test. Since the pooled value of v_e for the core densities accounted for only 5.5 percent of the total variation in the data, the additional materials variation introduced by the offset could not be great.

Materials and Testing Variability

A number of interesting observations can be drawn from Table 2. The most obvious involves the small to insignificant effect of transverse location. This result signifies that there was little or no consistent difference in the measured parameters as a function of transverse location. It was true not only for the pooled results, but for four of the five individual jobs as well. On only one job, District 3, did the free edge location exhibit a significant effect, having consistently higher air voids. However, this finding was not repeated for the District 3 relative compaction results. One must conclude that relative compaction and other properties studied here do not vary systematically as a function of transverse location. This is an important finding for designing sampling strategies. It means that transverse test locations randomly chosen provide unbiased estimates of materials quality.

As expected, longitudinal location was a significant factor for all parameters measured. It accounted for about half of the total variability in original densities. Also significant was the transverse/longitudinal interaction, accounting for nearly as much variability in the original core densities as the longitudinal effect; on two of the five jobs, it actually accounted for more. The value of v_{TL} is the real measure of transverse variability. The fact that it appears as an interaction with longitudinal location simply means that the transverse variability is not consistent from station to station.

The pooled values for v_L and v_{TL} are approximately equal for both the core and gage in-place densities. This means that there was, on average, as much variability in compaction across the mat at a single station as there was from station to station over the length of the job. The ratio, v_L/v_{TL}, ranged between 1.1 and 2.9 for three of the five jobs. For the District 5 job, v_L/v_{TL} equaled 11.2. This was the longest job, at 16.6 lane-km (10.3 lane-mi), but also the only one that drew material from two sources. The lowest value was 0.11 for the District 3 job. It was one of the shortest projects, at 5.3 lane-km

(3.3 lane-mi), and had the best longitudinal control. These observations are important because of their implications for lot size. For jobs using a single source, lot sizes much larger than now used may be practical.

The overall importance of the compaction operation can be inferred by examining the difference in total materials variability ($v_T + v_L + v_{TL}$) between the original and recompacted densities.

Variability attributable to the compaction operation should not affect the recompacted data, whereas the compactability of the material will. To be consistent, Method C was used throughout for this analysis. In overall terms, 67 percent of the original core density variability was removed as a result of the standardized recompaction. For three of the projects, this was well over 80 percent. Variability in the compaction operation, not the compactability of the material, was the major contributor to total density variability on these jobs. For the District 5 job, however, only 12 percent of the variability was removed by standardized recompaction. In this case, materials compactibility was more variable because multiple sources were involved.

Another indication of the importance of a consistent and well-coordinated compaction operation can be seen by examining the v_L and v_{TL} components of variance. The ratio of the variances of the original to recompacted densities for the longitudinal component is nearly 2:1. Contrast this with a ratio of nearly 10:1 for the transverse/longitudinal interaction. This implies that most variability in original densities attributable to compaction operations occurs in the transverse direction. Attempts to reduce this problem should focus on strategies that minimize random, transverse variability.

Test error variance was about 10 times greater for the nuclear gage than for the core specific gravities. This variance means that approximately 10 gage readings must be averaged to achieve the same precision as one core density result. The advantage of the gage over coring is its speed: more tests can be performed over a broader area, thereby better characterizing the larger locational sources of variability.

The sum of the materials variance components for the nuclear gage results was more than twice that of the cores. Yet both methods were run on the same material. The reason for this between-site variance could not be determined, but its magnitude was significant, equaling twice the total of the within- and between-gage variance. It is possible that gage drift contributed to the additional variability. Readings were taken over a 3- to 4-hr period, sufficient time for gage drift to become a problem. Unfortunately, only a beginning standard count was made, so it is impossible to ascertain the amount of drift that might have occurred. Another possible explanation is that the readings were influenced by the density of the underlying layers, adding to the overall variability of the locational factors.

Accuracy of the nuclear gage method is also an issue of concern. The gage densities were consistently lower than the core densities for all projects in the study; others have reported similar findings (*6,7*). On average, the district gage relative compaction results were 2.7 percent less than the core results (e.g., 92.3 versus 95 percent). This ranged from a low of 0.6 percent to a high of 4.1 percent for individual projects, a range consistent with the 1.5 to 5 percent range reported by Alexander (*1*). The care taken in drying the cores before testing should have minimized any residual moisture left by the coolant water used in the coring operation, often cited as a reason for high core densities. As Alexander reports, the surface voids in-filled by wax during the water displacement measurement can lead to overestimates of bulk densities of about the same magnitude as observed here.

The test error for asphalt content based on the hot extraction procedure contributed 39 percent of the total variation observed in the extraction data. At 0.19 and 0.30 percent, respectively, the test error and total standard deviations compare well with results reported by others (*8–10*). The other major contributor is the longitudinal component, v_L. Given the nature of asphalt content as a parameter controlled primarily at the plant, the significance of v_L is not unexpected. These results make it clear, however, that reflux extraction tests are too imprecise to use in any ERS involving pay adjustment factors. Unfortunately, reduction of test error through multiple test averaging is not practical given the complexity and cost of the test. The nuclear asphalt content gage, while achieving about the same degree of precision on a single sample, offers better potential for testing multiple samples quickly and efficiently. It should be the favored method for any ERS involving asphalt content.

A certain amount of variability in final lift thickness is to be expected to meet smoothness specifications, but the variability in these projects was surprisingly large. Longitudinal location explained about twice as much of the variation as did transverse location. The overall standard deviation of approximately 1 cm (0.03 ft) means that a planned thickness of 4.5 cm (0.15 ft) can vary anywhere from 2.5 to 6.5 cm (0.08 to 0.21 ft) over 95 percent of the job. The remaining 5 percent will deviate beyond these extremes. For the five projects studied, the overall average lift thickness of 5.2 cm (0.17 ft) exactly equaled the overall average design value. Job averages ranged from 15 percent below to 25 percent above the design value. These extremes are moderate, however, when compared with the within-job variation of ±45 percent. The lowest thickness recorded for one project was 1 cm (0.03 ft), 78 percent below the design thickness.

The aggregate grading for the extracted cores shows an unacceptably high test error component for the coarse sieves. This result is not surprising given the small volume of the core samples. In Figure 3, the percentage of total variation attributable to test error, longitudinal location (v_L), and transverse location ($v_T + v_{TL}$) are plotted as a function of sieve size. The relative contribution of test error reaches a minimum for the #16 sieve and then begins to increase again. In absolute terms, however, v_e continues to decrease. The longitudinal component is far more significant than the transverse components, at least for the #4 sieve and finer. For such fractions, longitudinal variance is anywhere from two to six times greater than

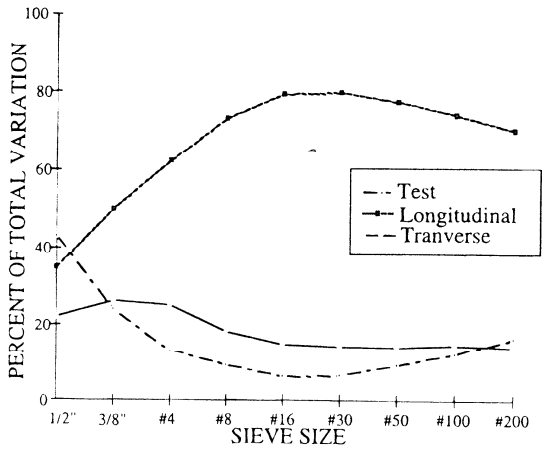

FIGURE 3 Relative contributions of variance components by sieve size.

transverse variance. For the coarser fractions, the values are close to equal. The fact that fine aggregate grading exhibits a relatively small variation in transverse location indicates that paver segregation problems are more important for coarse aggregates.

The high test errors for coarse sieves from the cores prompted a comparison with the grading and asphalt content results for the loose street samples collected by state inspectors during placement. Table 3 presents a summary of the job averages for both types of samples. Four of the five projects exhibited higher passing percentages for the core samples on the coarse sieves. The two sources for the District 5 project did not show this same pattern. Though not conclusive, the potential bias introduced by the coring process, in which some particles get smaller and none get larger, can be estimated roughly at anywhere from 0 to 7 percent additional passing coarse sieves for 10-cm (4-in.) cores. The implication of this finding for coring programs is that either larger cores should be taken or gradings and extractions should be based on loose samples collected from the uncompacted mat.

End Result Versus Method Specifications

From a previously published study, the relative compaction job average and between-lot standard deviation were available for 16 ERS jobs (2). Lots were approximately 454 T (500 tons) in size with the lot average based on 10 individual gage readings at random locations.

Figure 4 summarizes the results of the nuclear gage relative compaction for the 16 ERS projects and the 5 method specification jobs. The 95 percent target value is shown as a dashed line in the figure. A Mann-Whitney test of the job means between the two types of specifications easily rejects, at the 99 percent confidence level, the hypothesis that the means are from the same population (*11*). On average, the ERS jobs showed a gain in relative compaction of 3.1 percent over the method specification jobs (94.9 versus 91.8 percent). Approximately 47 percent of the results from the ERS jobs met or exceeded the 95 percent target, contrasted with only 9 percent for the method specification jobs.

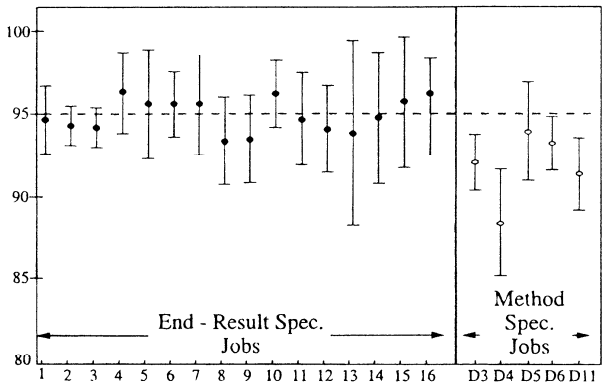

FIGURE 4 Mean and 95 percent confidence limits for nuclear gauge relative compaction results.

An opportunity for verifying this finding under a more controlled set of circumstances was available for the District 3 job. Immediately after this project was completed, the same contractor/supplier was awarded an adjacent project on the same route but with an ERS specification for relative compaction. Using the same mix design and essentially the same personnel and equipment, the contractor improved his average relative compaction from 91.8 to 94.6 percent, a gain of 2.8 percent relative compaction.

To make a direct comparison between lot to lot variability for the two types of jobs, the sum of the variance components for the method specification jobs was divided by 10 to represent the vari-

TABLE 3 Comparison of Test Results from Cores and Loose Samples

District	Sample Type	Grading (% Passing) 1/2"	#4	#30	#200	Asphalt Content (%)
3	Core	87	60	26	7	5.2
	Loose	81	53	21	4	---
4	Core	86	53	16	4	4.8
	Loose	84	51	16	4	4.7
5A	Core	83	49	23	4	5.4
	Loose	86	51	23	4	5.5
5B	Core	90	51	22	5	5.7
	Loose	90	52	21	5	6.3
6	Core	85	55	23	4	5.2
	Loose	78	52	28	4	4.9
11	Core	98	62	26	7	6.1
	Loose	97	59	24	6	6.6

ance of averages of 10. The adjusted values were used to construct the 95 percent confidence bands shown in Figure 4. Although average relative compaction clearly improved under the ERS approach, in most cases there was no similar improvement in job uniformity (note the approximately equal 95 percent confidence bands shown for both types of jobs). The District 3 matched comparison was an exception to this pattern, with total variance being reduced by 65 percent. Still, most contractors did not improve the control of their operations under the California ERS but simply worked harder to achieve the specified compaction. It is not surprising that few improvements in product uniformity were observed since the California ERS requires the contractor to achieve only a minimum average relative compaction of 95 percent with pay reductions enforced for any lots under 93 percent. Under an ERS that used a quality index or percentage defective approach, there would be much greater incentive for contractors to minimize materials variability.

The overall average void content for the five method specification jobs was 9.6 percent, ranging from 7.7 for the District 11 job to 12.2 for District 4 (Table 1). The standard deviation for all the jobs was relatively uniform at about 2 percent. For optimal performance, initial voids should not exceed 8 percent and, in service, not fall below 3 percent (*12*). Only 20 percent of the material placed in the method specification jobs fell within this zone. From the findings of this research, had these projects been performed under an ERS, this result would have increased to about 75 percent. By incorporating elements into the ERS that would encourage contractors to improve uniformity as well as average level, this percentage could be increased even further.

Lot Size

Significant economies in testing for asphalt concrete would be realized by going to larger lot sizes. The typical lot in California is 454 T (500 tons), approximately half a day's production. Under a method specification, this relatively small lot size makes sense because the inspector should be on the job continuously during the paving operation. Under an ERS, however, the contractor's personnel will be responsible for the day-to-day quality control. By increasing the lot size to 5000 T, a test crew would need to visit the job only once a week. This approach lends itself to regionalizing testing personnel and economies of scale.

An important assumption in the selection of a lot size is the homogeneity of the lot. The material contained in the lot needs to come from a relatively continuous operation, and the quality parameters should be distributed unimodally. The variable length of the five projects in this study offered an opportunity to examine the correlation of longitudinal variability to lot size for a variety of test parameters. A summary of the longitudinal variance components (shown as standard deviations) and project parameters is presented in Table 4.

For virtually all the results, the two shortest jobs rank lowest in terms of longitudinal variability. Only for the fine grading does this pattern break. The two longest jobs clearly have the highest v_L's for relative compaction, but a pattern is not obvious for the other tests. In terms of relative compaction based on the cores, the increase in longitudinal variability from the short jobs to the long jobs was approximately a third of the total variability. This is certainly a significant increase, but one that could be accommodated and even reduced under a well-designed ERS.

On the basis of this information, increasing the lot size for determining asphalt concrete density to 1 week's production is feasible from a statistical standpoint. Initially, risks will be higher as contractors learn to control the quality of their product. In the long run, the savings realized by a more efficient testing program and the improvements in product quality expected under an ERS will reduce these risks to acceptable levels.

CONCLUSIONS

On the basis of the five method specification projects studied, it can be stated with confidence that ERSs yield better compliance with asphalt concrete density requirements than method specifications. California and other states that have not already done so should adopt ERS as the standard method for specifying compaction of asphalt concrete. It is easy to imagine that the same principles of incentive and direct control at work in the successful application of ERS to compaction control can be extended to other properties, materials, and products as well.

Given the potential for improving job uniformity as well as average quality level, any ERS adopted should be based on a fraction defective, fraction within limits, or quality index approach (*13*). Since premature maintenance expenditures can be triggered by scattered, localized failures, achieving uniformity is just as important as controlling the average quality level of a product.

A number of related findings on materials variability were made as a by-product of this work. The ANOVA results point to areas in

TABLE 4 Relationship of Longitudinal Variance Component to Job Size and Haul Time

				Standard Deviation of the Longitudinal Variance Component						
	Duration of Work	Length	Haul Time	Relative Compaction (%)		Air Voids	Asphalt Content	Percent Passing		
District	(Days)	(Lane-km)	(Min.)	Cores	Gage	(%)	(%)	#4	#30	#200
6	1	2.90	10	0.99	1.06	1.10	0.05	1.59	1.17	0.94
3	2	5.31	30	0.48	0.99	0.53	0.09	1.31	1.12	0.47
11	3	5.63	15	1.38	2.27	1.43	0.39	7.34	3.70	0.34
4	4	13.52	15	1.34	3.29	1.63	0.14	1.96	1.01	0.20
5	5	16.58	60	1.38	2.80	1.39	0.15	2.70	2.16	0.55

which the largest potential for reduction in variability exists. In some cases, such as asphalt content, test precision should be improved. In others, either better plant control or better field control is indicated. For compaction, the results clearly show that efforts are best expended on improving the compaction operation by reducing random transverse variability.

The ANOVA results also revealed that larger lot sizes than are now used are feasible. Within the practical limitations of the type of job, lot size could be expanded tenfold to encompass an entire week's production. There are considerable benefits in terms of reduced staff and equipment inventory to be realized by state departments of transportation if larger lot sizes are implemented. The increases in risk to buyers and sellers as a result of slightly higher within-lot variability are not unreasonable.

ACKNOWLEDGMENTS

This work was conducted by Caltrans with funding from FHWA. The author is indebted to Caltrans staff members Fermin Barriga, Robert Cramer, Dawn Becky, Ken Iwasaki, Dick Wood, and student assistant Jorge Escobar for their support in collecting and analyzing the data.

REFERENCES

1. Alexander, M. L. *Control of AC Compaction Using an End-Result Specification.* Report 633189. California Department of Transportation, Sacramento, 1985.
2. Alexander, M. L. *Cost/Benefit Evaluation of End-Result Asphalt Concrete Compaction Specifications.* Report 643384. California Department of Transportation, Sacramento, 1988.
3. Hicks, C. R. *Fundamental Concepts in the Design of Experiments.* Holt, Rinehart and Winston, New York, 1964.
4. Kennedy, J. B., and A. M. Neville. *Basic Statistical Methods for Engineers and Scientists.* Harper and Row Publishers, New York, 1986.
5. Wall, F. J. *Statistical Data Analysis Handbook.* McGraw-Hill Book Company, New York, 1986.
6. Kennedy, T. W., M. Tahmoressi, and J. N. Anagnos. *A Summary of the Field Compaction of Asphalt Mixtures in Texas.* Report 317-2F. University of Texas at Austin, 1986.
7. Burati, J. L., Jr., and G. B. Elzoghbi. Correlation of Nuclear Density Results with Core Densities. In *Transportation Research Record 1126,* TRB, National Research Council, Washington, D.C., 1987, pp. 53–67.
8. Germain, J. J. *Evaluation of Microwave Oven and Nuclear Asphalt Content Gauge.* FHWA, U.S. Department of Transportation, 1987.
9. Christensen, D. W., and J. P. Tavris. *Nuclear Method Determination of Asphalt Content Corrected for Moisture in Bituminous Mixture.* Interim Status Report. Pennsylvania State University, University Park, 1988.
10. Stroup-Gardiner, M., D. E. Newcomb, and J. A. Epps. Precision of Methods for Determining Asphalt Cement Content. In *Transportation Research Record 1228,* TRB, National Research Council, Washington, D.C., 1989, pp. 156–167.
11. Bradley, J. V. *Distribution-Free Statistical Tests.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
12. Brown, E. R. Density of Asphalt Concrete—How Much Is Needed. In *Transportation Research Record 1282,* TRB, National Research Council, Washington, D.C., 1990, pp. 27–32.
13. Afferton, K. C., J. Freidenrich, and R. M. Weed. Managing Quality: Time for a National Policy. In *Transportation Research Record 1340,* TRB, National Research Council, Washington, D.C., 1992, pp. 3–39.

Publication of this paper sponsored by Committee on Management of Quality Assurance.