Using the Results of Contractor-Performed Tests in Quality Assurance

DETAILS

17 pages | | PAPERBACK
ISBN 978-0-309-43621-2 | DOI 10.17226/23135

AUTHORS

November 2007

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Subject Areas: IA Planning and Administration, IC Transportation Law,
IIB Pavement Design, Management, and Performance, IIIB Materials and Construction

Responsible Senior Program Officer: E. T. Harrigan

Research Results Digest 323

USING THE RESULTS OF CONTRACTOR-PERFORMED TESTS IN QUALITY ASSURANCE

This digest summarizes key findings from NCHRP Project 10-58(02), "Using Contractor-Performed Tests in Quality Assurance," conducted by Auburn University, Auburn, Alabama. The digest is an abridgement of the project final report that was authored by F. Parker, the principal investigator, and R. E. Turochy of Auburn University. Select chapters of the contractor's final report are available as *NCHRP Web-Only Document 115*.

CONTENTS

INTRODUCTION

The construction quality assurance process is comprised of process (quality) control, acceptance, and independent assurance procedures. Traditionally, contractors are responsible for their quality control, and state departments of transportation (DOTs) are responsible for acceptance and independent assurance. However, with the enactment of a federal regulation in 1995, commonly referred to as "23 CFR 637B," the roles of state DOTs and contractors have become less clear cut. Under certain conditions, 23 CFR 637B permits the use of contractor test results for acceptance. There is general agreement on the value of contractor quality (process) control. Issues may arise, however, when the results of contractor-performed tests are used in the acceptance process.

This research employed statistical procedures to evaluate whether state DOTs can effectively use contractor-performed test results in the quality-assurance process. The results of state DOT- and contractor-performed tests for hot mixed asphalt concrete (HMAC), portland cement concrete (PCC), and granular base course were collected and statistically compared. Field projects were selected to allow evaluation of as many as possible of the quality-assurance variables that might affect the comparisons.

The null hypothesis of this research was that the contractor-performed tests for use in the acceptance decision provide the same results as state DOT-performed tests. To test this hypothesis, contractor and state DOT results from six states were statistically compared to determine if differences between them in (1) variability and (2) proximity to target or limiting values were significant at $\alpha = 0.01$.

Comparison of Contractor-Performed Test Results with State DOT Test Results

HMAC test results were collected and analyzed from six state DOTs: Georgia, Florida, North Carolina, Kansas, California, and New Mexico. Details of verification and acceptance procedures for these six state DOTs are presented in Table 1.

The verification and acceptance procedures presented in Table 1 provide a range of details that might affect comparisons of contractor- and state DOT-performed tests.

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Table 1 Details of hot mixed asphalt concrete verification and acceptance procedures

State DOT	Properties	Contractor to DOT Testing Frequency	Verification Comparisons	Acceptance Method	Lot Size	Acceptance Data	Acceptance Criteria	Pay Factor Application
Georgia	AC, Gradation	4 to 1[2]	1 to 1	Adjust Pay	Days Production	Contractor	Absolute Deviation from Targets	Lowest[3] Pay
Florida	AC, VTM, Gradation, Mat Density	4 to 1 and 8 or 12 to 1	1 to 1	Adjust Pay	2000 or 4000 tons[1]	Contractor	PWL	Weighted Average
North Carolina	AC, VTM, VFA, Gradation, Mat Density	10 to 1 and 20 to 1	1 to 1	Adjust Pay[4]	Mix-Indefinite Mat Density-Days Production	Contractor[3] and DOT	Deviations from Target	Lowest[4] Mix Pay and Mat
Kansas	VTM, Mat Density	Mix 4 to 1 Mat 2 to 1	*F*- and *t*-Tests	Adjust Pay	Mix-3000[T] Mat Density–Days Production	Contractor	PWL	Both[5] Mix and Mat
California	AC, Gradation, Mat Density	10 to 1	*t*-Test and 1 to 1	Adjust Pay	Project Production	Contractor	PWL	Weighted Average
New Mexico	AC, VTM, Gradation, Mat Density	3 to 1	*F*- and *t*-Tests	Adjust Pay	Project Production	Contractor and DOT	PWL	Weighted Average

NOTES:
1. Contractor chooses 2,000 or 4,000 ton lots for acceptance.
2. Will vary based on production rate but data provided indicates about 4 to 1.
3. Mat density pay adjustments are included but are based on state DOT tests.
4. Pay adjustments (reduction only) applied independently for mix properties and mat density. Control charts with both contractor and DOT test results used to control mix production process and to decide when pay reductions applied. Mix pay reductions appear to be a last resort. Mat density pay computed for each lot.
5. Mix pay factor based on VTM and mat density pay factor applied independently.

Ratios of number of contractor to number of state DOT tests range from 2 to 1 to 20 to 1. Three state DOTs use simple 1 to 1 comparisons of test results to verify contractor-performed tests. More statistically robust comparisons of variances and means with *F*- and *t*-tests are used by the other state DOTs. Pay adjustments are applied by all six DOTs, but only as a last resort for mix properties by the North Carolina DOT. The North Carolina DOT acceptance procedure for HMAC mix properties is an accept/reject procedure based on control chart monitoring of both state- and contractor-performed tests. Lot size varies from 2000 tons or a day's production to the entire project production. Acceptance is based on verified contractor tests or combined DOT and verified contractor tests. When contractor-performed tests are not verified, acceptance is based on state DOT tests. Acceptance criteria include deviations from target values, absolute deviations from target values, or deviations from targets and variability with the percent within limits (PWL) method. The weighted average lot pay factor for all properties considered is applied by three state DOTs. The lowest pay factor from all properties considered or the pay factors for all properties considered are applied by the other three state DOTs.

Portland cement concrete pavement (PCCP) strength data were collected and analyzed from the Colorado DOT, and granular base course data were collected and analyzed from the FHWA-Western Federal Lands (FHWA-WFLHD). Details of verification and acceptance procedures are presented in Table 2.

Tests results generated during an entire construction season for a particular material were requested from the state DOTs listed in Tables 1 and 2. Some provided the requested data, some provided partial data from a construction season, and some provided limited data from several construction seasons. This collection resulted in a wide range in the size of data sets. Examples are presented below.

Examples

The North Carolina DOT provided HMAC tests for 735 mix designs from the 2004 construction year. This gave data sets with over 14,000 contractor mix tests, over 2,000 North Carolina DOT mix tests, over 20,000 contractor mat density tests, and over 6,000 North Carolina DOT mat density tests.

The Florida DOT provided HMAC tests from 98 selected projects constructed during the 2003 and 2004 construction years. This gave data sets with over 2,000 contractor mix tests, over 500 Florida DOT mix tests, over 6,000 contractor mat density tests, and over 1400 Florida DOT mat density tests.

The Colorado DOT provided portland cement concrete pavement (PCCP) flexural strength tests from three projects constructed in 2000, 2001, and 2003, respectively. The total data sets were comprised of 221 contractor tests and 61 Colorado DOT tests.

STATISTICAL ANALYSIS TECHNIQUES

Variability and proximity to target or limiting values (means) of contractor- and state DOT-performed tests were statistically compared. Variability, as measured with variance, was compared with *F*-tests. The proximity to target or limiting values, as measured with means of differences between test results and target or limiting values, were compared with *t*-tests. Means for contractor and state DOT tests from split samples were compared with paired *t*-tests.

Mean square deviation (*MSD*) provides a way to evaluate process control that considers both accuracy (proximity to target) and precision (variability) of the process. Mean square deviations for contractor and state tests were compared to determine which is the best indicator of material quality (process control). For the "nominal is best" (*NIB*) situation, where test results may be either larger or smaller than targets, the *MSD* is computed with

$$MSD_{NIB} = \frac{\sum_{i=1}^{n}\left(X_i - X_T\right)^2}{n}$$

where

X_i = test results,
X_T = target, and
n = number of measurements.

For large *n* values, this can be written as

$$MSD_{NIB} = s^2 + \left(\bar{X} - X_T\right)^2$$

where

s^2 = variance of tests and
$\bar{X}$ = mean of tests.

The variable used to combine tests with different target values is the difference between tests and target

Table 2 Details of portland cement concrete and granular base course verification and acceptance procedures

Agency	Material	Properties	Contractor to Agency Testing Frequency	Verification Comparisons	Acceptance Method	Lot Size	Acceptance Data	Acceptance Criteria	Pay Factor Application
Colorado DOT	PCC Pavement	Flexural Strength	4 to 1	*F*- and *t*-Tests	Adjust Pay	Project Production	Contractor	PWL	—
FHWA-WFLHD	Granular Base	Gradation, LL, PI, % Fractured Particles and SE/P200	10 to 1 after first 3 for a project[1]	*F*- and *t*-Tests	Adjust Pay	Project Production	Contractor	PWL	Lowest[2]

1. Data provided indicates this results in an average testing frequency ratio of about 3 to 1.
2. Pay adjustment for density also included but based only on contractor-performed tests. FHWA-WFLHD witnesses density testing.

values. Therefore, the most desirable value is always zero and the equation for MSD_{NIB} reduces to

$$MSD_{NIB} = s^2 + \left(\bar{\Delta}\right)^2$$

where $\bar{\Delta}$ = mean of difference between tests and target values. Smaller MSD_{NIB} values for manufacturing processes mean better control. When comparing MSD_{NIB} values for two sets of test results for the same process, smaller MSD_{NIB} indicate more precise tests with closer conformity to target values.

In some cases, for example, a specification for in-place mat density with a minimum acceptable mat density requirement, the "largest is best" (*LIB*) situation is applicable when computing mean square deviations. The LIB mean square deviation (MSD_{LIB}) can be approximated with the equation

$$MSD_{LIB} \approx \frac{1}{\left(\bar{X}\right)^2}\left[1 + \frac{3s^2}{\left(\bar{X}\right)^2}\right]$$

where

$\bar{X}$ = mean of measurements and
s^2 = variance of measurements.

All statistical comparisons were made at a 1% level of significance ($\alpha = 0.01$), which provides a stronger determination of differences than the more widely used 5% level of significance ($\alpha = 0.05$).

All data provided by a state DOT for a particular material were combined for analyses. Some properties have target values that vary by project or job mix formula, for example asphalt content of HMAC. Therefore, it was necessary to subtract target values from measurements ($\Delta = X - X_T$) to produce a variable that could be combined for all mixes or projects.

In addition, data from small projects ($n_{DOT} < 6$ or projects less than six state DOT test results) were eliminated to produce reduced data sets that combined data from larger projects. These were analyzed to determine if project size might affect statistical comparisons. In addition, comparisons and analyses were conducted for larger projects ($n_{DOT} \geq 6$ or projects with more than six state DOT test results) on a project-by-project basis, i.e., variability and means for contractor and state DOT were compared for each project or, for the North Carolina DOT, for each job mix formula.

ANALYSIS OF GEORGIA DOT HMAC DATA

HMAC test results obtained by Georgia DOT during the 2003 construction year were analyzed. Measured properties included gradation (percent passing 1-in., ¾-in., ½-in., ⅜-in., #4, #8, #50, and #200 sieves) and asphalt content measured with either the vacuum solvent extraction or ignition methods. Quality control (QC) samples were taken and tested for each 500-ton sublot. Results from these tests are used for lot acceptance if verified by the results of Georgia DOT comparison tests, for which comparison samples are split from the contractor's QC samples for one of every 10 lots with results compared one to one.

The paired *t*-test was used to compare the means of results between the contractor QC and Georgia DOT comparison tests. Significant differences in deviation from target values were found for only four of eight sieves. However, the differences were significant for three of the four sieves used for pay adjustment computation. Mean deviations from target values were larger for Georgia DOT tests for five of eight sieves. However, for the four sieves used for pay adjustment computation, deviations for Georgia DOT tests were always larger (as noted above, significantly so for three of four sieves). The deviations from target asphalt contents were not significantly different, but the deviations for Georgia DOT tests were larger.

The *F*-test was used to compare variances between the contractor QC and Georgia DOT comparison gradation tests. The variances were significantly different for four of eight sieves, including those of three of the four sieves used in pay adjustment computation. Numerically, the variances for Georgia DOT gradation tests were larger for seven of eight sieves. The variance for the Georgia DOT asphalt content tests was significantly larger than the variance for contractor tests.

Comparisons between contractor QC and Georgia DOT QA test results were made with the *t*-test, as these test results are from independent samples and one to one comparisons with the paired *t*-test were not appropriate. Variances were compared with *F*-tests. The contractor QC test results compared with Georgia DOT comparison test results discussed above are a subset of the total contractor QC test results data set.

These comparisons indicate that, except for the percent passing the 1- and ¾-in. sieves, the variances of Georgia DOT test results are significantly larger than variances of contractor test results. However, no significant differences in the means of any of the test results were found. The Georgia DOT means for the percent passing the #4 sieve used for pay are larger, but the contractor mean for asphalt content is larger.

Mean square deviation (*MSD*) provides a method for considering both accuracy (proximity to target) and precision (variability) in evaluating measurements. MSD_{NIB} were calculated for the set of contractor QC tests, the set of Georgia DOT comparison tests, and the set of Georgia DOT QA tests. Values for contractor QC tests are smallest for all properties except percent passing the ¾-in. sieve. This implies that contractor tests are consistently more precise and closer to target values.

The MSD_{NIB} for Georgia DOT QA tests are closer to MSD_{NIB} for contractor QC tests for percents passing #4 sieve and asphalt content. The MSD_{NIB} for Georgia DOT comparison test results are closer to MSD_{NIB} for contractor QC tests for percent passing #4 sieve. This was surprising because Georgia DOT comparison and contractor QC samples are split samples and test results are directly compared one to one. It is reasonable to assume that this would promote similarities. Georgia DOT QA test results are from independent samples and results are compared to acceptance criteria. It may be that this more direct relationship with the acceptance process is the reason contractor QC and Georgia DOT QA tests are more comparable than contractor QC and Georgia DOT comparison tests.

Variance and, therefore, measurement precision dominates the computation of MSD_{NIB}. Target values are zero and means for the differences from targets, when squared, are small. Except for percent passing the #50 sieve for Georgia DOT comparison and QA tests, comparisons of variances would provide the same relative rankings as MSD_{NIB}.

Project-by-project comparisons of Georgia DOT and contractor tests for asphalt content and percent passing the ½-in. and #200 sieves were also conducted in an effort to quantify the numbers of projects where (1) there are significant differences between Georgia DOT and contractor means and variances and (2) Georgia DOT means and variances are largest. These comparisons generally confirm trends indicated by comparisons of combined tests; namely, that the variability of Georgia DOT tests is likely larger than the variability of contractor tests, but that means of Georgia DOT tests are less likely larger than means of contractor tests.

Except for asphalt content, these project-by-project analyses indicate no particular tendency for Georgia DOT or contractor tests to be closer to target values. Nor is there any strong tendency for means of differences from targets to be significant but, when differences are significant, the Georgia DOT means are likely larger.

The Georgia DOT means are larger for 68 (60%) of the projects. These means are significantly different for eight (7%) of the projects and significantly larger for six (5%) projects.

The Georgia DOT variances are larger for 77 projects (68%). These variances are significantly different for 12 projects (10%); for 10 projects (9%), the Georgia DOT variances are larger.

ANALYSIS OF FLORIDA DOT HMAC DATA

Florida DOT provided HMAC test results from 98 projects constructed during 2003 and 2004. These test results were described by the state as "an excellent sampling of the types of mixture properties that are used and a good sampling of the contractors that conduct FDOT work."

Test results included gradation (percents passing ¾-in., ½-in., ⅜-in., #4, #8, #16, #30, #50, #100, and #200 sieves), asphalt content, maximum specific gravity (G_{mm}), bulk specific gravity of laboratory compacted samples (G_{mb}), air voids and VMA (computed with G_{mm} and G_{mb}), and mat density (% G_{mm}, core bulk specific gravity as a percentage of G_{mm}). Percent passing the #8 sieve, percent passing the #200 sieve, asphalt content, air voids, and mat density (%G_{mm}) were used in the PWL system to compute lot composite pay factors. Asphalt content and gradation were determined with the ignition oven method. Mat density was measured with 6-in. cores.

Lot size was 2,000 or 4,000 tons (contractor choice) divided into either four 500-ton or four 1,000-ton sublots. Contractors tested one mix sample and five cores per sublot. Florida DOT conducted two types of sampling and testing: (1) verification and (2) independent sample verification testing (ISVT). Florida DOT verification tests and contractor QC tests were conducted on split samples.

Variances and means of differences from target values were compared between contractor QC and Florida DOT verification test results from split samples for data from (1) all projects and (2) large projects (those with at least six Florida DOT test results [$n_{FDOT} \geq 6$]). Data from the large projects were also compared on a project by project basis.

The comparisons for all projects and for large projects were very consistent. Variances of contractor and Florida DOT test results were generally significantly different. Exceptions were for percent passing the #16, #30, #50, and #100 sieves. Proximity to target values was consistently not significantly different, although those values for mat density (%G_{mm}) approached statistically significant differences.

In all cases, variances of Florida DOT test results were greater than variances of contractor test results.

Except for percent passing the #50 sieve and asphalt content among large projects, the mean differences indicated contractor test results closer to target values. For the cases of all and large project comparisons, mean differences from target asphalt contents were quite small.

The target values used for VMA are minimum acceptable values. The negative mean differences of about 0.5 % indicated lower than desirable VMA values. Contractor VMA measurements were greater and, therefore, closer to minimum acceptable values.

Comparisons between contractor QC and Florida DOT ISVT test results from independent samples for all and for large projects were reasonably consistent. Variances of Florida DOT ISVT and contractor test results are significantly different, except for (1) percent passing the #50 sieve and (2) percent passing the #200 sieve for large projects. Mean values were generally not significantly different. Important exceptions were air voids and mat density (%G_{mm}) where means were significantly different. Mean values for percent passing the #4 and #8 sieves were also significantly different for test results from all projects.

For all cases, variances of Florida DOT test results were greater than variances of contractor test results.

Contractor gradation test results were closer to target values than Florida DOT test results, except for percent passing the ½-in. sieve. For asphalt content, Florida DOT test results were closer to targets. These differences for asphalt content are, however, quite small and are consistent with Florida verification test results. The VMA comparisons indicate more favorable Florida DOT ISVT test results; namely, greater test results relative to minimum acceptable values. This is the opposite of indications from comparisons with Florida DOT verification test results where contractor test results were more favorable, relative to minimum acceptable values.

For air voids, the mean differences indicate Florida DOT ISVT test results are significantly closer to the 4% target value than contractor test results. This is the opposite of comparisons with Florida DOT verification test results where contractor test results were closer to the target, but not significantly closer.

For mat density (%G_{mm}), the mean differences indicate contractor test results were significantly closer to target values than Florida DOT ISVT test results. The Florida verification test result comparisons also indicated contractor tests results closer to targets, but not significantly closer.

The third comparison was between means of paired contractor QC and FDOT verification test results. Paired *t*-tests for data from all projects and from large projects, respectively, were conducted, including comparisons of maximum mix specific gravity (G_{mm}), laboratory bulk specific gravity (G_{mb}), and core bulk specific gravity (G_{mb}), for which means were reported rather than means of differences between test results and targets.

The comparisons for gradation indicate significant differences in percent passing for all except the ½-in. and ⅜-in. sieves for large projects. These results were quite different from the previous comparisons of *t*-test results where none of the differences were significant. Since the magnitudes of the mean differences are similar in both comparisons, the inconsistencies may be attributed to the paired *t*-test being somewhat more discerning than the *t*-test.

The comparisons of paired percent asphalt content, air voids, and VMA test results were the same as the previous unpaired comparisons; that is, differences in means are not significant. Likewise, differences in G_{mm} and the G_{mb} of laboratory compacted samples were not significant.

Unlike the unpaired data, however, the differences for the paired percent G_{mm} test results were significant. This was surprising as were the significant differences for G_{mb} of cores. The same cores were tested by Florida DOT and contractors and the results used to compute %G_{mm}. Pairing test results from the same cores should make significant differences unlikely, but this was not the case. An analysis of the magnitude of mean differences of %G_{mm} found that the contractor results for compaction were much closer to target densities than any other test results.

Project-by-project comparisons for large projects (i.e., projects where there were six or more Florida DOT test results) were conducted to quantify the numbers of projects with (1) significant differences between Florida DOT and contractor means and variances and (2) greater Florida DOT means and variances. These comparisons generally confirmed trends found in comparisons of combined test results and are summarized as follows:

- Differences from target values of Florida DOT test results tend to be greater than contractor test results.
- Variances of Florida DOT test results are greater than the variances of contractor test results.
- Variances of project test results are more likely to be significantly different than mean differences.
- When mean differences from target values are significantly different, mean differences for Florida DOT test results are likely greater.
- When variances are significantly different, variances for Florida DOT test results are likely greater.

A final analysis compared values of MSD_{NIB} computed with means and variances. MSD_{NIB} for VMA are not included because Florida DOT specification contained minimum acceptable requirements. The values of MSD_{NIB} confirmed trends indicated by comparisons of mean differences and variances: contractor test results are more accurate, relative to target values, and more precise (less variable) than Florida DOT test results. The MSD_{NIB} for contractor test results are all smaller than Florida DOT test results from split (verification) and independent (ISVT) samples.

ANALYSIS OF NORTH CAROLINA DOT HMAC DATA

North Carolina DOT (NCDOT) provided test results for HMAC produced and placed during the 2004 construction year. HMAC production is managed by job mix formula (JMF) while compaction is managed by project; therefore, there is a disconnect between test results for HMAC properties and mat properties. Mat densities were obtained in a format such that sorting and, therefore, analysis was convenient only by JMF.

Comparisons of HMAC Properties

Test results for HMAC properties were obtained for 735 mix designs. These were combined into a data set of all JMFs and then sorted into a reduced data set comprised of JMFs where there was six or more NCDOT test results. Proximity to targets and variances of contractor and NCDOT test results were compared for the combined and reduced data sets. Comparisons were also conducted for each JMF with six or more NCDOT test results.

Test results included gradation (percent passing 1-in., ¾-in., ½-in., ¾-in., #4, #8, and #200 sieves), asphalt content, air voids, VMA, VFA, and % G_{mm} @ N_i in the gyratory compactor. The ignition oven method was used for asphalt content and gradation, except that the contractor could request an alternative method for asphalt content.

The lot size for contractor QC sampling and testing was 750 tons, but mix acceptance was not on a per lot basis. For mat compaction, a lot was a day's production. NCDOT conducts two types of mix sampling and testing: QA and verification. QA tests are conducted with contractor QC on split samples; verification tests are conducted on independent samples.

Comparisons were made between contractor QC and NCDOT QA test results for all 735 JMFs. Except for VMA, the variances of NCDOT and contractor test results were statistically significantly different. For all properties, NCDOT variances were greater.

Significant differences for means were not as consistent. Six of twelve properties had statistically significant different means. Except for VMA and %G_{mm} @ N_i, the means of differences from target values indicated contractor test results were closer to target values than NCDOT test results. The specification requirement for VMA is a minimum acceptable value and for %G_{mm} @ N_i is a maximum acceptable value. The means of differences indicate more favorable contractor test results for both VMA and % G_{mm} @ N_i.

The results of the analysis of the reduced data set confirmed these results, with a few differences for specific comparisons.

Comparisons of contractor QC and NCDOT QA test results with the paired *t*-test yielded statistically significant differences for means of all properties. Except for % passing the 1-in. sieve, contractor test results were either closer to target values or, for VMA and %G_{mm} @ N_i, more favorable relative to NCDOT specification requirements.

Comparisons of paired contractor QC and NCDOT QA test results in the reduced data set indicated more consistent statistically significant differences than comparisons of unpaired test results. Only the comparisons for percent passing the ¾-in. and 1-in. sieves were not significantly different.

The reduced dataset was analyzed by comparing JMF statistics. The analyses were similar to those project-by-project comparisons conducted with Florida and Georgia DOT test results. These JMF by JMF comparisons generally confirmed trends indicated by comparisons of combined test results, and may be summarized as follows:

- Differences from target values of NCDOT test results tend to be greater than contractor test results. Percentage differences are equal to 50% for passing ½-in. sieve and greater than 50% for all other properties.
- Variances of NCDOT test results are generally larger than the variances of contractor test results. The average percentage of JMFs with variances greater for NCDOT than contractors is 64% for the six properties used for acceptance. For the six properties not used for acceptance, the average is 54%.
- When mean differences from target values are significantly different, mean differences for NCDOT test results are likely larger (90 of 107 JMFs or 84%). The average percentage of JMFs with greater means for NCDOT than contractors is 13% for the six properties used for acceptance and 4% for the six properties not used for acceptance.
- When variances are significantly different, variances for NCDOT are likely larger (80 of 97 JMFs or 82%).

Similar comparisons were made between contractor QC and NCDOT verification test results from all 735 JMFs. These test results are from independent samples. Except for $\%G_{mm}$ @ N_i, the variances of NCDOT and contractor test results are statistically significantly different. For all properties, NCDOT variances are greater. The comparisons of variances for test results from independent samples are quite similar to the previous comparisons for test results from split samples.

Significant differences for means were not as consistent. In total, only three of twelve means (25%) are significantly different compared to six of twelve (50%) for the split samples discussed above. Contractor means of differences are smaller for all properties except VMA and $\%G_{mm}$ @ N_i. However, the mean differences for these properties also indicate more favorable contractor test results.

The MSD_{NIB} for contractor QC tests of all properties were smaller indicating the best process control (material properties). The MSD_{NIB} for NCDOT QA tests were largest for three properties and largest for seven properties for the North Carolina verification tests.

Mat Density Comparisons

Mat compaction is managed by project and acceptance is by lot where a lot is a day's production; contractors may choose testing with nuclear gauges or cores. Contractors conduct five tests at equal intervals in each 2,000-ft test section. NCDOT conducts a retest (same locations) in 10% of the test sections and conducts verification tests (independent locations) in 5% of the test sections. Results are reported as the average of the five tests.

Combined nuclear gauge testing for 141 JMFs was analyzed. The comparisons of variances and means are consistent for both NCDOT retest and verification results and for all large JMFs ($n_{NCDOT} \geq 6$). Variances of NCDOT nuclear gauge mat density tests are significantly larger than contractor tests.

When nuclear gauge test results were examined on a JMF-by-JMF basis, NCDOT JMF mean differences from the 92% G_{mm} minimum indicate lower in-place densities. In cases where the JMF mean differences between NCDOT and contractors are significantly different, the NCDOT values are likely smaller. When differences in variances are statistically significant, NCDOT variances are always larger.

As noted previously, contractors may also choose core testing for control and acceptance of mat compaction. Contractors take and test one core in each 2,000-ft test section. The NCDOT conducts retest (same cores) and test comparison cores (taken adjacent to contractor QC core locations) in 10% of the test sections. In addition, the NCDOT tests one verification core from an independent location in 5% of the test sections.

Analyses of combined core testing for 585 JMFs yielded consistent comparisons for NCDOT retest, comparison, and verification test results for the cases of (1) all and (2) large JMFs ($n_{NCDOT} \geq 6$):

- Variances of NCDOT core mat density tests are significantly greater than contractor tests.
- Mean differences of NCDOT core mat density tests from the 92% G_{mm} minimum target are significantly different from contractor tests, and NCDOT tests indicate poorer compaction.

The comparisons for core mat density tests are similar to those for nuclear gauge tests; namely, variances and means are significantly different. The mean differences from the 92% G_{mm} minimum target are also similar in magnitude. However, the variances for nuclear gauge and core tests are numerically different, with core variances consistently greater. Although there may be differences in the variability of nuclear gauge and core testing, some portions of the observed differences are likely due to sample size, since one core is taken per 2,000-linear foot test section, compared to five nuclear gauge tests run per test section, with the average recorded as a test result. Since the acceptance procedure is the same for either type testing, it is surprising that contractors chose the core option (larger variance) about twice as often as the nuclear gauge option (contractor $n_{CORE} = 20{,}282$ and $n_{NG} = 9{,}011$).

When core test results were examined on a JMF-by-JMF basis, NCDOT JMF mean differences from the 92% G_{mm} minimum indicated lower achieved densities. As with nuclear gauge tests, there were a surprising number of JMFs where contractor tests indicated densities exceeding the 92% G_{mm} minimum but NCDOT tests indicated densities less than the 92% G_{mm} minimum. When JMF mean differences are significantly different, the NCDOT values were always smaller.

NCDOT test variances are generally larger when differences in variances between NCDOT and contractor results are significant. An exception is the contractor QC and NCDOT retest core comparison where the NCDOT variances are greater for only four of 49 JMFs (8%). Since the same cores were tested by both agencies, the significant differences in mean differences from targets or variance were surprising.

Since NCDOT specifications have a minimum acceptable mat density requirement of 92% of G_{mm}, the "largest is best" situation is applicable when computing mean square deviations. To compute the MSD_{LIB} for mat density tests, the statistics (s^2 and $\bar{\Delta}$) for nuclear gauge and core tests of all JMFs were used. Means were computed by adding the minimum mat density requirement (92% of G_{mm}) to the mean deviations

$$\left(\bar{X} = \bar{\Delta} + 92\right).$$

The contractor MSD_{LIB} are smallest for both nuclear gauge and core tests of mat density indicating the best process control (mat compaction). The MSD_{LIB} for the several NCDOT tests are all relatively similar. Comparable values for nuclear gauge and core tests are close and are an indication that the means dominate the computations since the variances of core tests are considerably larger than variances of nuclear gauge tests.

ANALYSIS OF KANSAS DOT HMAC DATA

Test results from 49 projects constructed in Kansas during the 2003 season were analyzed. Properties compared were theoretical maximum mix specific gravity (G_{mm}), air void content of laboratory compacted specimens, and mat density. Mat density was typically measured with nuclear gages but could also be measured with cores. Gradation and asphalt content are measured by both contractors and Kansas DOT (KSDOT), but only for process control. Gradation and asphalt content test results are not archived by the KSDOT and, therefore, were not available for analysis.

A lot for measuring HMAC properties is 3,000 tons divided into four 750-ton sublots. Contractors take one QC sample per sublot and the KSDOT takes one independent verification sample for each lot, yielding a sampling ratio of four to one. A lot for mat density is a day's production divided into five sublots. Contractors make two and the KSDOT one independent mat density measurement for each sublot for a sampling ratio of two to one.

KSDOT has no target mat density but uses the PWL system for computing pay adjustments with a lower specification limit (LSL). To combine data from multiple projects with different LSL, the following variable was defined:

$$\Delta = X - \text{LSL}$$

where

X = 90 % for shoulder paving
= 91 % for mainline paving 2 in. thick and less
= 92 % for mainline paving thicker than 2 in.

Variances and means of KSDOT and contractor tests for combined data from all projects were compared. Variances for KSDOT tests are significantly greater for all comparisons. Means are significantly different for mat density (%G_{mm}) for the combined data and for both thin (≤2 in.) and thick (>2 in.) mainline paving.

The KSDOT mean of differences from the air voids target value is the greater. Contractor means of differences from mat density LSLs are larger and indicate better mat compaction.

The KSDOT and contractor data were sorted into a reduced data set for projects with $n_{KDOT} \geq 6$. In comparisons of variances and means for air voids and $\%G_{mm}$ for this reduced data set, KSDOT variances are significantly greater than contractor variances. The means of air voids are not significantly different but the contractor mean difference from lower compaction specification limits is significantly greater than the KSDOT mean difference.

Project-by-project comparisons include data for theoretical maximum mix specific gravity (G_{mm}). These comparisons indicate that differences from target air voids and G_{mm} test results are likely larger for KSDOT but that differences from mat density LSLs are likely larger for contractors. Only project deviations from mat density LSLs were likely significantly different; contractor project deviations were likely larger. Project variances for KSDOT tests were likely larger. When variances were significantly different, KSDOT variances were always largest.

MSD_{NIB} was most appropriate for air voids; MSD_{LIB} was applicable to mat density. The *MSD* for contractor tests were always smaller indicating better process control (material quality).

ANALYSIS OF CALTRANS HMAC DATA

HMAC test results from 149 projects constructed from 1996 to 2005 were provided by Caltrans. Caltrans' quality assurance procedures use both mix properties and mat density for acceptance, but only test results for mix properties were provided. Test results included asphalt content and gradation (percent passing ¾- or ½-in., ⅜-in., #4, #8, #30, and #200 sieves).

Caltrans defines a lot for acceptance as the entire project production. For contractor QC sampling and testing a sublot is 500 tons; Caltrans samples and tests for verification at a frequency not less than 10% of the contractor QC frequency. Caltrans samples independently for mix properties. Verification requires acceptable comparison of means with the *t*-test ($\alpha = 0.01$) and with numerical allowable testing differences.

Comparisons were made between contractor QC and Caltrans tests for all 149 projects. The variances for all seven properties were significantly different; the variances of Caltrans tests were always larger. The mean differences from target values for four of the seven properties were significantly different. Except for the percent passing the #30 sieve, the Caltrans mean differences from target values were the larger for these four properties. The mean differences from targets were not significantly different for three other properties, but the Caltrans mean differences from target values for these three properties were larger.

A reduced data set for large ($n_{CAL} \geq 6$) projects was created from the total data set. Comparisons of project means and variances for these larger projects provided the following trends:

- Differences from target values of Caltrans test results tend to be larger than differences from target values for contractor test results.
- Variances of Caltrans test results are larger than variances of contractor test results.
- Except for the percent passing the #4 sieve, when mean differences from target values are significantly different, mean differences for Caltrans test results are likely larger.
- When variances are significantly different, variances of Caltrans test results are generally larger.

A final comparison was made between MSD_{NIB} of Caltrans and contractor tests. The MSD_{NIB} for contractor tests are smaller, indicating better process control (material quality).

ANALYSIS OF NEW MEXICO DOT HMAC DATA

Limited HMAC data were provided by the New Mexico DOT (NMDOT). These data included results from three projects with seven mixes. Results from project analyses, rather than test results, were provided. These results included target values, statistics ($\bar{X}$ and s), and pay adjustments computed with both NMDOT and contractor statistics. Depending on the details of the project and mix, measured HMAC properties may include asphalt content, mat density, air voids, nominal maximum aggregate size, and percent passing the #4, #8, #10, #16, #30, #40, #50, and #200 sieve sizes.

NMDOT accepts HMAC on a lot-by-lot basis where a lot is defined as the entire project production for a particular mix design. Contractor QC and NMDOT acceptance test results are compared ($\alpha = 0.01$) with *F*- and *t*-tests as they are obtained. Verification requires that both variance and mean are not significantly different.

For the seven mixes, only six of 60 (10%) of the standard deviations were significantly different and contractor standard deviations were smaller for five of six cases. Seven of 60 (12%) of the means were significantly different but contractor means were closer to target values in only two of seven cases. Similarities between NMDOT and contractor test results may arise from the verification process where *F*- and *t*-tests are used to compare variability and accuracy as results are accumulated for the entire project mix design production.

If the standard deviations and means of differences from target values for common properties from New Mexico are compared with statistics from other states, some interesting findings emerge. Except for the standard deviation of contractor-tested air void contents in Kansas, the standard deviations for the NMDOT tests (both DOT and contractor) are always the smallest. For asphalt content, the variability is considerably smaller than that of any other state. There is also no consistent indication that the variability of contractor test results is smaller than the variability of NMDOT tests or that differences are significant.

The standard deviations for both NMDOT and contractor asphalt content tests are 0.116, which is about two and a half times smaller than the standard deviations of asphalt content tests for any other state. However, it should be noted that the asphalt content standard deviations are for data from only three New Mexico projects and they are not unlike standard deviations for individual projects in other states. The similarity of the contractor and NMDOT data may be due to the limited size of the database compared to the other states.

The means for New Mexico asphalt contents are in line with the other states and indicate that, on average, test results are quite close to target values. The NMDOT air voids content results are much closer to the four percent target value than test results for any of the other four states, except for the Florida DOT independent sample verification test (ISVT) results.

Mean mat densities for FDOT and NMDOT reflect differences between measured and target values. Mat density means for KDOT and NCDOT reflect differences between measured and lower specification limits or minimum acceptable values. This mean was much closer to the target than the mean for any other type Florida DOT test.

To summarize, the statistics for the limited NMDOT data are quite different from the statistics in the other five states studied. The NMDOT and contractor test results appear more similar in variability and accuracy. The reasons for the observed differences and similarities are not known. Possible factors include the verification and acceptance system that defines a lot as the entire project mix production, the accumulation and comparison of DOT and contractor test results with *F*- and *t*-tests, the combining of DOT and contractor test results to make acceptance decisions, or some combination of these factors. While the Caltrans system is similar to that of the NMDOT, the variances for Caltrans asphalt content and percent passing the #200 sieve are larger. Variances among Caltrans tests are more like variances for Georgia, Florida, North Carolina, and Kansas DOT tests than the NMDOT tests.

ANALYSIS OF COLORADO DOT PORTLAND CEMENT CONCRETE PAVEMENT DATA

Flexural strength test results from three portland cement concrete pavement (PCCP) projects were provided by the Colorado DOT. Contractors could choose between acceptance processes that used either 28-day flexural or compressive strengths. With the compressive strength process, contractor test results were used only for quality control and DOT test results were used for acceptance. There was no required comparison of compressive strength test results. With the flexural strength process, contractor and Colorado DOT test results were compared with *F*- and *t*-tests ($\alpha = 0.05$). If the contractor tests were verified, they were used for acceptance. Comparisons had to indicate no significant difference for both variances and means for verification.

A lot was defined as the entire project production of a process, where consistent materials, mix design and construction method were used. Contractors fabricated and tested a set of three beams per 2,500 m^2 of pavement or a minimum of one set of three beams per day. The Colorado DOT independently fabricated and tested a set of three beams per 10,000 m^2 of pavement. A test result was the average flexural strength of three beams.

Contractor and Colorado DOT flexural strength test results were compared. Data were combined from the three projects; the analysis variable was the difference between test results and lower specification limit flexural strength ($\Delta = x - x_L$). The comparisons indicated no significant differences ($\alpha = 0.01$) between Colorado DOT and contractor flexural strength test results.

There were also limited PCCP data available from several other states. Comparisons of portland cement concrete (PCC) compressive strength test results from Kentucky and Alabama, conducted at α = 0.05 significance level, indicated no significant differences in variances or means for structural PCC. There were also test results for paving PCC from Kentucky. Here the comparisons indicated there was no significant difference in means of the Kentucky paving PCC compressive strength, but that there was a significant difference in variances. These limited comparisons suggest that, if there are significant differences between contractor and DOT test results, it is more likely these will be differences in variability.

The comparisons for the Colorado, Alabama, and Kentucky PCC strength data suggest no particular tendency for contractor tests to be less variable or more favorable (higher strengths) as was generally found for HMAC.

ANALYSIS OF FHWA–WESTERN FEDERAL LANDS HIGHWAY DIVISION AGGREGATE COURSE DATA

The FHWA-Western Federal Lands Highway Division (FHWA-WFLHD) provided test results from 23 aggregate course construction projects. The projects involved several types of aggregate courses. Each type of aggregate course had some combination of properties for pay factor computation, drawn from LL, PI, SE/P200, percent fractured particles, and percents passing the 1-in., ¾-in., ½-in., ⅜-in., #4, #10, #40, and #200 sieves. For these properties both contractor and FHWA-WFLHD testing of split samples were required. FHWA-WFLHD tests were used to verify contractor tests with *t*- or paired *t*-tests at a 1% significance level. If verified, contractor tests were used to compute lot pay factors with the PWL method. Pavement layers have compaction requirements, but layer compaction is accepted or rejected based on contractor density tests.

A lot is the entire project production for a particular type of aggregate course. Contractors took and tested one sample per 1,000 tons of aggregate placed. The FHWA-WFLHD tested a split sample from the first three project samples and at least 10% of the remaining project samples. The data provided for the 23 projects indicated an average contractor to FHWA-WFLHD testing ratio of about three to one. However, the ratio for a particular project depended on the project quantity. The variable used for the comparisons was the difference between test results and either target values, maximum specification values, or minimum specification values. Target, maximum, and minimum values were subtracted from test results.

Comparisons for the entire data sets of FHWA-WFLHD and contractor tests suggest that differences in variability are not as extensive as those for HMAC but the differences in means are somewhat more extensive. The variabilities of only five of 12 properties were significantly different but, for these five, the variability of FHWA-WFLHD tests was larger for four properties. Overall, the variabilities of eight of 12 FHWA-WFLHD test properties were larger. This observation of larger agency test variability is consistent with observations for HMAC tests.

The procedure FHWA-WFLHD used to establish gradation targets affected consideration of the comparisons of means. Gradation targets were set as the average of contractor tests, provided the average was within specification allowable limits. For example, if the allowable range for percent passing a sieve is 20 to 30% and the average for contractor tests is 27%, the target value would be 27%. This was the case for most of the 23 projects and accounts for the low gradation mean deviations for contractor tests. As a result, comparisons of the magnitude of FHWA-WFLHD and contractor mean deviations from gradation targets are not meaningful.

The gradation means were significantly different for only two of seven sieves. For the remaining properties, the FHWA-WFLHD test means were significantly different for three of four properties. The contractor means were more favorable, relative to specification limits for these three properties. The means for SE/P200 were not significantly different and the FHWA-WFLHD mean was slightly more favorable relative to its minimum specified values.

Means of contractor and FHWA-WFLHD test results from split samples were compared with paired *t*-tests. The comparisons of gradation means were the same as those for the entire contractor data set described above. Means for only two sieves (¾ and ½ in.) were significantly different. The comparisons of paired tests for the remaining samples were the same as the comparisons for all data, except for PI. The means for paired PI tests were not significantly different.

The comparisons of mean differences from target values for granular base are similar to comparisons for HMAC. Means are not consistently significantly different but, when they are significantly different, contractor tests are likely more favorable (LL, PI, and % fractured particles).

Project-by-project comparisons were made for the nine properties, LL, PI, percent fractured particles, and percents passing the ½-in., ⅜-in., #4, #10,

#40, and #200 sieves. Percentages passing the 1-in. and 3/4-in. sieves and the ratio of the sand equivalent and % passing the #200 sieve (SE/P200) were omitted because individual project data were insufficient for meaningful comparisons. Projects were included that had five or more FHWA-WFLHD tests. Previous project analyses defined a large project as one with six or more agency tests. However, these data included a number of projects with five FHWA-WFLHD tests and the inclusion of these projects greatly expanded the database.

These analyses indicated contractor gradation tests were consistently closer to target values. However, this finding is influenced by the designation of the project target percent passing as the average of contractor tests, provided this average is within specification tolerances. The results indicate it is unlikely that the gradation means are significantly different between FHWA-WFLHD.

No particular tendency for FHWA-WFLHD or contractor gradation variances to be larger, except for percent passing the #4 sieve, was identified. Nor were variances of gradation tests likely to be significantly different, but, if they were, the variances of FHWA-WFLHD gradation tests were always larger.

EFFECT OF DIFFERENCES IN TEST RESULTS ON ACCEPTANCE OUTCOMES

An assessment of the hypothetical effects of the observed differences between state DOT and contractor tests was conducted on the evaluations of acceptance outcomes using HMAC statistics for Georgia, Florida, Kansas, North Carolina, and California. Statistics for contractor and state test results were applied to acceptance procedures to compute theoretical outcomes. The outcomes were compared and the differences computed. Table 3 presents the results of these computations. Outcomes computed with contractor test results were more favorable than those computed with state test results; this is expected since contractor test results are, typically, closer to targets and significantly less variable.

Table 3 Comparison of acceptance outcomes with DOT and contractor test results

	% Greater Chance of				
Property	**PF<100% with Georgia DOT Statistics**	**Exceeding Spec. Limits with Florida DOT Statistics**	**Exceeding Spec. Limits with Kansas DOT Statistics**	**Exceeding Min. Spec. Limits with North Carolina DOT Statistics**	**Exceeding Spec. Limits with Caltrans Statistics**
% Asphalt	0.8	6.0	—	—	7.8
% Passing 1/2″ Sieve	0.1	—	—	—	2.7
% Passing 3/8″ Sieve	0.3	—	—	—	5.4
% Passing #4 Sieve	2.7	—	—	—	1.5
% Passing #8 Sieve	12.3	7.6	—	—	5.6
% Passing #30 Sieve	—	—	—	—	5.9
% Passing #200 Sieve	—	5.7	—	—	4.7
Void Content	—	12.5	13.9	—	—
Mat Density (% G_{mm})	—	3.6 Coarse Mix 4.2 Fine Mix	10.3	12.6* 9.1**	—

*Contractor and NCDOT Retest—Nuclear Gage
**Contractor and NCDOT Comparison—Cores

ANALYSIS OF ASPHALT TECHNICIAN SURVEYS

A speculative cause for differences between state DOT and contractor-performed tests could be a motivation of technicians working for these organizations to provide test results perceived to provide more favorable acceptance outcomes for their employers. To examine this possibility, a survey was administered to contractor, consultant, and state DOT asphalt technicians in two settings: (1) in class to 21 technicians in a course sponsored by Florida DOT and (2) by mail to approximately 500 NICET certified asphalt technicians. While neither survey meets the accepted criteria for a rigorous scientific survey, they did provide some interesting findings.

The survey consisted of six questions:

1. My employer is
   - ☐ a state department of transportation,
   - ☐ a contractor,
   - ☐ a consultant, or
   - ☐ other.
2. I am involved in sampling and testing to control the production and placement of construction materials and/or the acceptance of these materials. Acceptance may be pass/fail or involve adjustments to bid prices.
   - ☐ Yes—continue
   - ☐ No—stop
3. Have you ever felt pressure to produce test results, or to retest, to give more favorable control or acceptance outcomes?
   - ☐ Yes—continue ☐ No—go to Question 5
4. Was the pressure you felt to produce test results that would give more favorable outcomes
   - ☐ self-induced—you just felt you should, or
   - ☐ due to specific reasons/instructions/comments from supervisors?
5. How easy/difficult would it be to manipulate test results to achieve more favorable outcomes?

   Easy Difficult

   1 2 3 4 5
6. Please rank, from 1 (most effective) to 5 (least effective), the following techniques for preventing manipulation of test results.
   a. sampling and testing of split samples for comparison
   b. sampling and testing of independent samples for comparison
   c. occasional observation of sampling and testing procedures
   d. use contractor-performed tests for process control and state DOT-performed tests only for acceptance
   e. periodic (weekly or monthly) audit and comparison of contractor and state DOT test results by an independent organization

Table 4 summarizes the results of the two surveys. Of the total of 161 respondents to the two surveys, a majority (97 or 60%) are employed by consultants; 25 (16%) and 14 (7%) were employed by contractors and state DOTs, respectively. Of the surveyed technicians employed by consultants, contractors, and state DOTs, overwhelming majorities (94%, 92%, and 93%, respectively) were involved in testing of construction materials for purposes of control, acceptance, or both.

Question 3 attempted to assess whether technicians have ever felt pressure to produce test results that gave more favorable outcomes to their employers. Substantial percentages of the technicians employed by consultants, contractors, and state DOTs (66%, 57%, and 23%, respectively) answered this question in the affirmative. It is interesting that the greatest percentage of positive responses came from the technicians working for consultant firms that presumably are more likely to be under contract to the state DOTs than paving contractors in control and acceptance activities. Overall, both contractor and state DOT technicians affirmatively answered this question in lower percentages (57% and 23%) than the 60% found for all technicians regardless of their type of employer. It must be cautioned that no follow-up question assessed whether the perceived pressure led to actual manipulation of test results.

Question 4 was only asked of those who had answered YES to Question 3. Fifty percent indicated that such pressure came from reasons, comments, or instructions from their supervisors. The survey did not attempt to further define the term *supervisor,* whether, for example, (1) it was limited to first-line supervisors or (2) it represented a perception of what the organization desired. Another 23% indicated that the pressure did not arise from supervisory directions but rather was self-induced through some undefined cause. A further 12% indicated that the pressure arose from a confluence of self-inducement and supervisory influence.

Table 4 Responses to the Florida DOT course and NICET asphalt technician surveys

Question 1		Question 2		Question 3		Question 4*		Question 5		Question 6 Method—Rank	
FLDOT	**NICET**	**FLDOT**	**NICET**	**FLDOT**	**NICET**	**FLDOT**	**NICET**	**FLDOT**	**NICET**	**FLDOT**	**NICET**
Consultants 2 (11%)	Consultants 95 (66%)	Yes-1 (50%) No-1 (50%)	Yes-90 (95%) No-5 (5%)	Yes-0 (0%) No-1 (100%)	Yes-61 (67%) No-30 (33%)	—	A-33 (54%) B-11 (18%) C-8 (13%) D-7 (11%) E-1 (2%) F-1 (2%)	1.00	1.28	a—1 b—2 c—3 d—5 e—4	a—1 b—2 c—3 d—5 e—4
Contractors 8 (44%)	Contractors 17 (12%)	Yes-8 (100%) No-0 (0%)	Yes-15 (88%) No-2 (12%)	Yes-3 (38%) No-5 (62%)	Yes-10 (67%) No-5 (33%)	A-0 (0%) B-3 (100%)	A-6 (60%) B-3 (30%) C-1 (10%)	3.12	1.37	a—2 b—1 c—4 d—5 e—3	a—2 b—1 c—4 d—5 e—3
FLDOT 7 (39%)	DOT 7 (5%)	Yes-7 (100%) No-0 (0%)	Yes-6 (86%) No-1 (14%)	Yes-1 (14%) No-6 (86%)	Yes-2 (33%) No-4 (67%)	A-1 (100%) B-0 (0%)	A-1 (50%) B-1 (50%)	1.50	1.26	a—2 b—3 c—4 d—1 e—5	a—1 b—5 c—2 d—4 e—3
Other 1 (6%)	Other 24 (17%)	Yes-0 (0%) No-1 (100%)	Yes-23 (96%) No-1 (4%)	—	Yes-15 (65%) No-8 (35%)	—	A-5 (33%) B-3 (20%) C-2 (13%) D-4 (27%) F-1 (7%)	—	1.24	—	a—1 b—2 c—3 d—5 e—4
All 18 (100%)	All 143 (100%)	Yes-16 (89%) No-2 (11%)	Yes-134 (94%) No-9 (6%)	Yes-4 (25%) No-12 (75%)	Yes-88 (65%) No-47 (35%)	A-1 (25%) B-3 (75%)	A-45 (52%) B-18 (21%) C-11 (12%) D-11 (12%) E-1 (1%) F-2 (2%)	2.33	1.27	a—1 b—2 c—4 d—3 e—5	a—1 b—2 c—3 d—5 e—4

*A: Supervisor; B: Self; C: Supervisor and Self; D: Contractor; E: Clients; F: No Response

The responses to Question 5 suggest that the technicians surveyed, regardless of their employer, consider that it would be relatively easy to manipulate test results to give more favorable results. Their responses to Question 6, in turn, indicate that the best way to forestall such manipulation by technicians employed by consultants, contractors, or state DOTs is the sampling and testing of split or independent samples for purposes of comparison.

SUMMARY OF KEY FINDINGS

The null hypothesis of this research was that the contractor-performed tests for use in the acceptance decision provide the same results as state DOT-performed tests. To test this hypothesis, contractor and DOT results from six states were statistically compared to determine if differences between them in (1) variability and (2) proximity to target or limiting values were significant at $\alpha = 0.01$.

For HMAC, the differences in means and variances found between the contractor and state DOT results are commonly significant. In general, the variability of state DOT quality assurance test results is larger than the variability of contractor quality control test results. Such differences might arise in part from differences (1) in the number of specimens commonly tested by contractor and state agency technicians and (2) in the time between sampling and testing of specimens often found between contractors and state agencies.

The statistical test results for PCC pavement and aggregate course construction are favorable toward pooling of contractor and state DOT results, although this finding is based on a smaller sampling of data than for HMAC. While there are no compelling reasons at this time not to use contractor-performed PCC tests for quality assurance, additional analyses would be prudent before this practice is generally adopted for PCC pavement and aggregate course construction.

THE NATIONAL ACADEMIES™

Advisers to the Nation on Science, Engineering, and Medicine

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people's lives worldwide.

www.national-academies.org

TRB

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001