TRANSPORTATION RESEARCH RECORD

No. 1491

Materials and Construction

Construction: Specifications and Management

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1995

Transportation Research Record 1491
ISSN 0361-1981
ISBN 0-309-06153-9
Price: $23.00

Subscriber Category
IIIB materials and construction

Printed in the United States of America

Sponsorship of Transportation Research Record 1491

GROUP 2—DESIGN AND CONSTRUCTION OF TRANSPORTATION FACILITIES
Chairman: Michael G. Katona, US Air Force Armstrong Lab

Construction Section
Chairman: Donn E. Hancher, University of Kentucky

Committee on Flexible Pavement Construction and Rehabilitation
Chairman: Robert B. McGennis, The Asphalt Institute
Timothy B. Aschenbrener, Fouad M. Bayomy, Elton R. Brown, Wayne Byard, John A. D'Angelo, Dale S. Decker, Cindy K. Estakhri, Colin A. Franco, Peter F. Hueppi, Charles S. Hughes, James S. Moulthrop, James G. Murfee, Frazier Parker Jr., Harold R. Paul, Freddy L. Roberts, Mark D. Sargent, James A. Scherocman, Richard J. Schreck, Ottmar W. Schutz, Robert M. Smith, Richard R. Stander Sr., Shelley M. Stoffels, Maghsoud Tahmoressi, Timothy W. Vollor

Committee on Management of Quality Assurance
Chairman: Charles S. Hughes
Secretary: Ronald J. Cominsky, Brent Rauhut Engineering Inc.
AbdulRahman Al-Azzam, Gabrielle E. Bedewi, Paul E. Benson, James L. Burati Jr., Michael I. Darter, Dale S. Decker, Robert P. Elliott, Ralph D. Ellis Jr., Foad Farid, Jim Gee, Dimitrios Goulias, Rosemary M. Ingram, Charles A. Kline, Peter A. Kopac, Rita B. Leahy, M. Lee Powell III, Orrin Riley, Gerald J. Rohrbach, Carlos E. Rosenberger, James M. Shilstone Sr., Clinton E. Solberg, Peter A. Spellerberg, Donald R. Tuggle, Richard M. Weed

Committee on Construction Management
Chairman: Arunprakash M. Shirole, New York State Department of Transportation
Yvan J. Beliveau, Eugene K. Blasko, Martin L. Cawley, Gary B. Chullino, Ralph D. Ellis Jr., Foad Farid, Richard A. M. Felsinger, George C. Gibson, Gary A. Gilmore, Rosemary M. Ingram, Ram Krishna, G. John Kurgan, Hosin Lee, Robert B. Newman, M. Lee Powell III, Christopher E. Reseigh, Jerry M. Riggsbee, Orrin Riley, James E. Rowings Jr., H. Randolph Thomas Jr., Thomas R. Warne, Connie Yew, Gordon D. Zwillenberg

Committee on Fabrication and Inspection of Metal Structures
Chairman: Fred R. Beckmann, American Institute of Steel Construction
John M. Barsom, Bart S. Bergendahl, Thomas A. Calzone, Nicholas M. Engelmann, John W. Fisher, Donald J. Flemming, Karl H. Frank, Al Ghorbanpoor, Ian D. Harris, Charles J. Hellier, Theodore Hopwood II, Henry G. Justus, Abba G. Lichtenstein, Philip A. Malachowski, David L. McQuaid, Dennis R. Mertz, Tom W. Neal, Robert L. Nickerson, M. Noyszewski, David W. Prine, Raymond W. Stieve, Godci K. Tehrani, John P. Weisner

Transportation Research Board Staff
Robert E. Spicher, Director, Technical Activities
Frederick D. Hejl, Engineer of Materials and Construction
Nancy A. Ackerman, Director, Reports and Editorial Services

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1994.

Transportation Research Record 1491

Contents

Foreword

The papers in this volume, which deal with various facets of construction, should be of interest to state and local construction, materials, and research engineers as well as contractors and material producers.

The first six papers address construction quality management specifications. Benson compares end result specification with method specification in California. Aurilio and Raymond discuss the development of an end result specification for pavement compaction in Ontario. Weed describes a PC program to generate operating characteristic curves for statistical construction specifications. Ksaibati et al. summarize and analyze pavement construction smoothness specifications used by U.S. state departments of transportation. Weed relates the development of air voids specifications for bituminous concrete in New Jersey. Hossain and Parcells report on the development and implementation of smoothness specifications in Kansas.

The last two papers address construction management. Abdul-Malak et al. present a decision support system framework for contractors to use when evaluating the feasibility of adopting advanced construction technologies. Roddis et al. describe and discuss an operational expert system for bridge fabrication.

PART 1

Construction Quality Management Specifications

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

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

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.

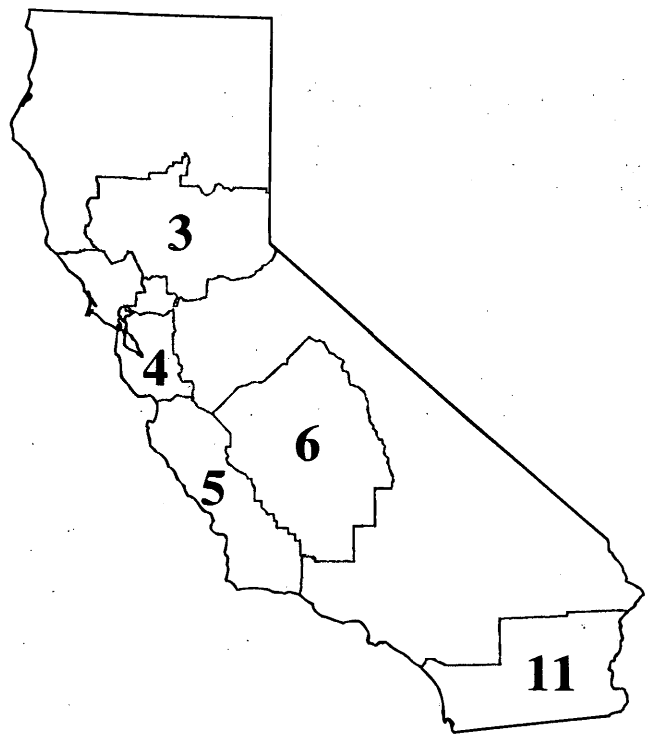


FIGURE 1 Participating districts.

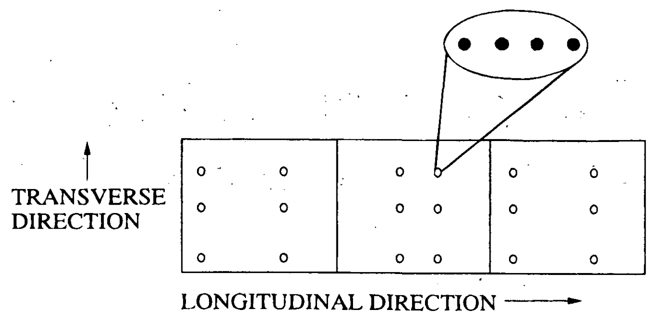


FIGURE 2 Sampling plan schematic.

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.

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_G 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

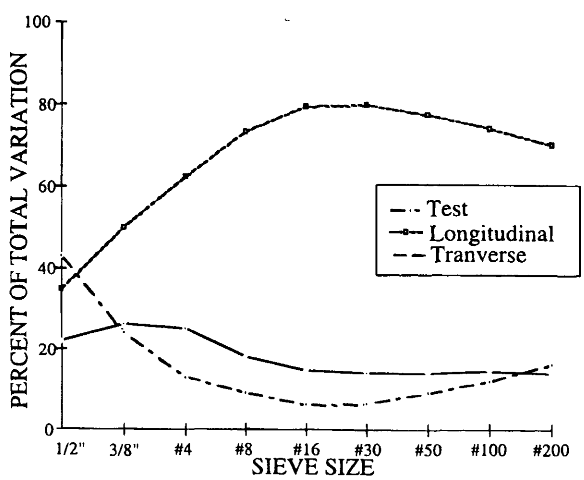


FIGURE 3 Relative contributions of variance components by sieve size.

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.

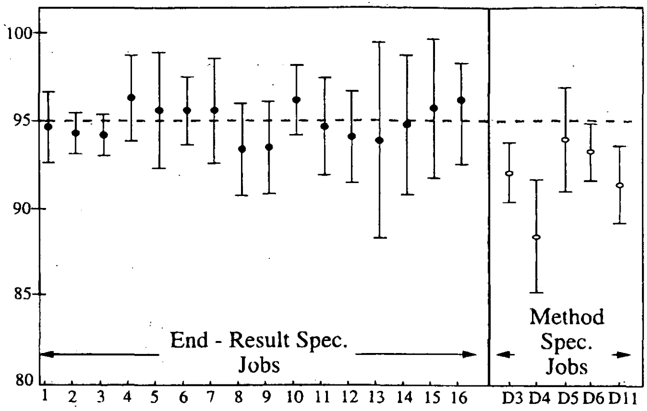


FIGURE 4 Mean and 95 percent confidence limits for nuclear gauge relative compaction results.

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.

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.

Development of End Result Specification for Pavement Compaction

V. AURILIO AND C. RAYMOND

In 1992 the Ontario Ministry of Transportation (MTO) developed a new, statistically based end result specification (ERS) for the acceptance of hot-mix asphalt. As part of the ERS phase-in plan for hot mix, the specification for pavement compaction was introduced to the industry with the intent for full implementation by 1995. The acceptance procedure employs a percent-within-limits specification using the lot mean and standard deviation to ensure that the desired compaction is achieved. Using a life-cycle cost analysis, the appropriate payment factors were calculated on the basis of the expected life of the final product. Operating characteristic curves were developed to analyze buyers' and sellers' risks and to evaluate the expected payment factors for the acceptance plan. Simulations were carried out to assess the effects of the proposed price adjustment system. The system provides a bonus for consistent compaction that exceeds a specified quality level and an adjustment in contract price for work that does not comply with the specification.

Over the past 15 years, the Ontario Ministry of Transportation (MTO) has been moving toward replacing many of its existing method specifications with statistically based end result specifications (ERSs). In 1987 MTO began its phase-in of ERS for hot mix, and in 1991 the first specification incorporating price adjustments for deficient material was implemented for the acceptance of hot mix based on asphalt cement content and full aggregate gradation. ERSs for other highway construction materials such as unbound aggregates, Portland cement concrete, and bridge-deck waterproofing were already in place.

At the start of this process MTO recognized the need to move slowly into ERS to allow stakeholders a chance to understand this new concept. A phase-in plan was developed in consultation with industry for the development and implementation of future ERSs. On the basis of the plan tabled in 1992, the pavement compaction specification is scheduled for implementation on 10 to 15 contracts in 1994 and on 50 percent of the new contracts in 1995. In addition to this formal implementation, MTO is offering contractors the option to mutually agree to have hot mix accepted under the new special provision without price reductions provided they perform process control and take responsibility for rejectable material.

The new specification was developed in fall 1992 and was presented to the Ontario Road Builders Association in March 1993. The specification is based on a percent-within-limits (PWL) philosophy. This approach is different from the one used for the first ERS implemented in 1991, which is based on a variability-known acceptance plan that assumed a constant or known variability. The PWL system was chosen primarily because it is considered to be a better indicator of quality. The use of a PWL system was recommended in a report (*1*) prepared on behalf of MTO to review ministry ERSs.

As part of the development and implementation of the new specification, the ministry and road builders agreed to carry out field simulations during fall 1993 to allow MTO staff and contractors the opportunity to gain experience with the specification (and the PWL acceptance system). The simulation would also provide an opportunity for industry to develop a quality-control (QC) plan and to identify any problems with the proposed specifications.

In Ontario QC is the responsibility of the contractor; although it is not the intent of MTO to specify QC requirements, the importance of good QC cannot be understated.

This paper includes a limited statistical analysis of historical data and describes the development of the new specification as well as a simulation study to illustrate how the acceptance plan works. Data from the 1992 construction season were used to simulate a distribution of the estimated PWL. Operating characteristic (OC) curves are shown based on the PWL distribution and continuous price adjustment schedule.

BACKGROUND

Compaction is considered to be one of the most important factors that affect the ultimate performance of hot-mix asphalt (HMA) pavement (*2*). Pavement compaction is critical for the development of internal strength and good durability properties. The literature indicates that for each 1 percent increase in air voids above 7 percent, there is a 10 percent decrease in the service life of the pavement (*3*).

The current specification for pavement compaction classifies rollers on the basis of roller width, roller diameter, and static mass and requires a contractor to use a specified combination of rollers depending on the rate of hot-mix production. The pavement compaction requirements (amended by special provision for MTO contracts) specify that the lot average shall be equal to or greater than 92 percent of the theoretical maximum relative density (MRD) with no single test value less than 90 percent. An additional requirement of this special provision is that the pavement density is corrected on the basis of actual core thickness. The correction factor C adjusts the pavement density by +0.1 percent for every 1-mm deviation below 40 mm; if the core thickness is less than 25 mm, the core is not used for compaction calculations and a replacement core is taken. The correction factor C accounts for the effect of thin lifts on compaction; it originated from a previous compaction specification that stipulated the level of compaction based on lift thickness.

$$\text{Percentage compaction} = \frac{\text{BRD}}{\text{MRD}} \times 100 + C$$

V. Aurilio, John Emery Geotechnical Engineering Limited, No. 1, 109 Woodbine Downs Boulevard, Etobicoke, Ontario, M9W 6Y1 Canada. C. Raymond, Ontario Ministry of Transportation, 1201 Wilson Avenue, Downsview, Ontario, M3M 1J8 Canada.

where

BRD = bulk relative density of individual core,
MRD = theoretical maximum relative density determined for lot, and
C = thickness correction factor (0.1 percent for each whole millimeter that pavement course thickness is less than 40 mm).

Under the existing specification, lots are based on a day's production and acceptance is based on core samples; three random cores are taken when the day's production is less than or equal to 1500 T, and one core is obtained for each 500 T when the production is greater than 1500 T. This system does not price-adjust material that does not meet specification.

A review of the 1991 data indicated that not all mix types were attaining the same level of compaction and that some mix types had a large percentage of material below the specification limit. To improve the overall quality of the pavements being constructed, MTO elected to proceed with an ERS for pavement compaction.

DATA ANALYSIS

The pavement compaction data from all 1992 contracts were compiled and analyzed by contract, mix type, and region. A statistical analysis was conducted to determine the mean, standard deviation, and coefficient of variation for each population tested. Frequency histograms were plotted to verify that the populations are distributed normally.

A summary of the analysis for pavement compaction is presented in Table 1. The 1992 data were compared with a more limited study performed on 1991 data, which indicated essentially similar trends in both construction seasons. Typical frequency histograms for some of the mix types are shown in Figure 1.

The histograms plotted for each mix type confirmed that the populations are approximately normally distributed. The data analysis also shows that there are significant differences in the pavement compaction being attained for the various mix types. However, from the data analyzed it is unclear whether this variance is due to construction (i.e., improper compaction) or to mix characteristic (i.e., gradation, aggregate type, etc.). The most noticeable difference observed was for the compaction attained for DFC and HDBC mixes. These mixes are premium mixes incorporating 100 percent crushed aggregates and are used on high-volume roadways in Ontario.

The population mean for all mixes was 93.0 percent, with a standard deviation of the lot means of 2.0 percent. The pooled standard deviation was found to be 1.6 percent; the coefficient of variation (or measure of relative dispersion) ranged from 1.1 percent (for RHM) to 2.4 percent (for HL 8). The overall coefficient of variation using the pooled standard deviation was 1.7 percent.

To study the effect of the correction factor applied to cores less than 40 mm, the data were analyzed to determine the pavement densities for each mix type corrected and uncorrected. The analysis shows that the correction factor was applied on approximately 20 percent of the lots tested. However, most of these corrections were for minor deviations in thickness; 50 to 60 percent of the lots receiving a correction were corrected by only 0.1 percent. The data also indicate that the average and standard deviations for each mix type were virtually the same for the uncorrected data versus corrected data versus all data excluding the corrected values.

ACCEPTANCE PLAN

There are two commonly used acceptance plans (involving inspection by variables) for evaluating hot-mix quality characteristics. They are referred to in AASHTO R9-91 (Acceptance Sampling Plans for Highway Construction) as "variability known" and "variability unknown." These methods evaluate the acceptability of the material on the basis of mean and variability measured by testing. The variability-known acceptance plan assumes that the variability is known and constant. This type of plan evaluates the lot mean on the basis of acceptance criteria developed using an assumed (or known) variability for the lot. The plan then separately evaluates the lot variability to ensure that it is less than the assumed (or specified) value.

Acceptance may be determined using either the mean and range method or the mean and standard deviation method. The standard deviation method is normally used and is recommended by AASHTO, mainly because all the samples are used to measure variability rather the range method, which uses only the highest and the lowest values of a lot.

The variability-unknown type of acceptance plan assumes the lot variability to be unknown. The PWL method estimates the normal distribution of the material on the basis of the mean and standard

TABLE 1 Analysis of 1992 Compaction Data

Mix Type	Lot Mean (%)	Std. Dev. of Lot Means	Coefficient of Variation	Minimum Lot Mean	Maximum Lot Mean	No. of Lots Analyzed
DFC	90.9	1.78	2.0	85.6	96.6	193
HDBC	91.6	1.65	1.8	87.6	96.6	324
HL1	93.9	1.54	1.6	89.7	98.3	94
HL3	92.8	1.46	1.6	87.0	95.8	78
HL4	93.7	1.57	1.7	88.1	98.9	741
HL8	93.7	2.24	2.4	88.3	97.6	90
MDBC	94.7	1.30	1.4	91.7	96.9	47
RHM	93.6	1.07	1.1	90.1	96.1	130
Total	93.0	1.97	2.1	85.6	98.9	1697

[a]*Dense Friction Course,* [b]*Heavy Duty Binder Course, Hot Laid,* [c]*Medium Duty Binder Course,* [d]*Recycled Hot Mix*[e]
NOTE: Mix types as follows: DFC =

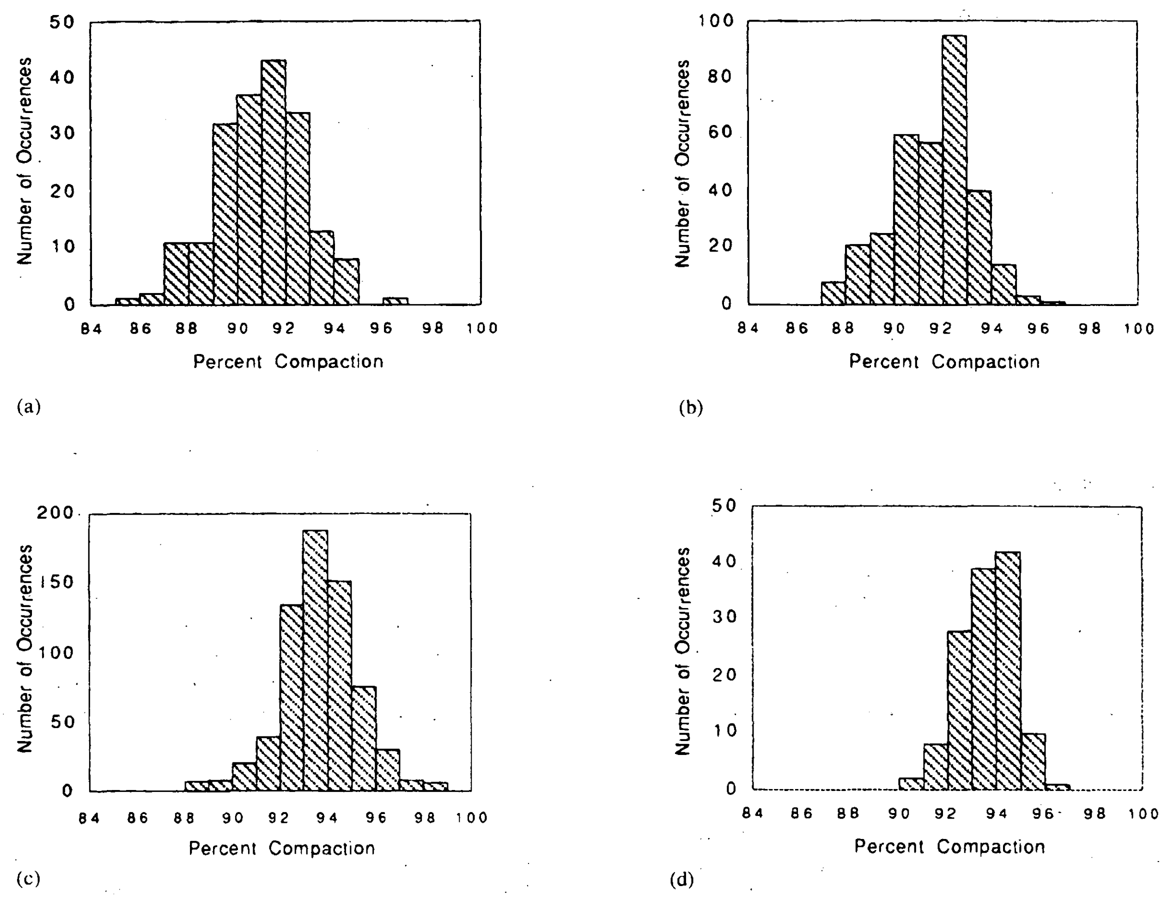


FIGURE 1 Frequency histograms for several mix types: *a*, DFC; *b*, HDBC; *c*, HL-4; *d*, RHM.

deviation of the test values. The distribution is then used to determine an estimate of the percentage of material within a lower or upper limit. The major advantage of this method is that the mean and variability (standard deviation) are used together in the same "equation" to estimate the quality of the material.

The acceptance plan developed for pavement compaction is based on a PWL principle. This method was selected for several reasons. It is widely accepted that PWL specifications are more efficient and beneficial for the contractor and the owner. Generally, this system provides a better estimate of the lot quality and is considered to be more effective. The estimate of PWL is unbiased and will most likely lead to fewer disputes about material quality. This system encourages uniformity of the end product, thereby improving the overall quality of the pavement (*4*). Last, because the standard deviation is a better measure of variability, fewer samples are required than when using the lot range. AASHTO R9–91 states that a range plan requires 12 samples to provide the same estimate of variability as a standard deviation plan using 10 samples.

SAMPLING PLAN

Several factors should be considered in determining the sampling frequency and lot size. The number of samples taken for a lot should be sufficient to ensure that the testing accurately represents the lot. Using a small number of samples will result in a high risk of incorrectly accepting unsatisfactory work (buyer's risk), a high risk of incorrectly rejecting good work (seller's risk), or both. From the buyer's perspective, the quantity of testing must also be practical to ensure that the cost is not excessive and that the testing can be carried out in a timely manner with the available resources. The lot size must be large enough to justify the expense of testing. However, if the lot is too large, the consequences associated with unacceptable material may become too severe. Another concern with a large lot is that the material is not uniform. This could occur from a change in the contractor's process or from other factors such as a major change in environmental conditions, which can affect compaction.

Lot sizes typically are based on either 1 day's production or a specific quantity. An advantage of decisioning a lot size on a daily basis is that environmental conditions and operational characteristics (i.e., rolling pattern or roller operator) are likely to remain more or less the same. These conditions and characteristics can deviate more when the lot size is based on a specified quantity, especially when production is slow or is stopped in the middle of a lot. However, with proper process control a contractor should be able to produce a uniform product.

To simplify the administration of accepting hot mix, it was decided that the lot sizes for pavement compaction would parallel the lot sizes specified for the current ERS for acceptance of asphalt cement content and aggregate gradation. Under this system a lot is normally defined as 2000 T of HMA with four sublots of equal portions. From the OC curves generated, a sampling frequency of six samples per lot (one per 333-T sublot) was chosen for the pavement compaction acceptance plan.

OC CURVES AND RISK ANALYSIS

The analysis of risk is considered to be an essential procedure when developing any ERS. By knowing the risks involved a contractor can establish a quality level that normally will guarantee full payment. Likewise, the owner can with some level of confidence ensure that product meets specification. The most common way to analyze risk is by developing OC curves. These curves generally relate the probability of acceptance or expected payment with a specific level of quality (i.e., PWL).

PRICE ADJUSTMENT SCHEDULE

Adjusted pay schedules are common with most ERSs. Price reductions normally are used to deal with materials that do not entirely meet specification but are not considered to be so substandard that removal or repair is required. To determine the appropriate pay adjustments, the design life of the pavement is compared with an expected life for the pavement (as-built) discounted over the life cycle of the pavement. This method is considered suitable provided a quality-versus-performance relationship can be established (*5*).

CALCULATION OF APPROPRIATE PAY FACTOR

The appropriate pay factors were determined using a life-cycle costing analysis with the model shown later. This analysis takes into account the original cost of hot mix (M_c) and allows for two resurfacings within the life span. The design life of the original pavement and each subsequent overlay is 10 years. This is typical of a design analysis performed by MTO. Inflation and interest rates are assumed to be 3.0 and 7.0 percent, respectively.

The appropriate pay factor is based on the present-worth cost of construction plus the cost of rescheduling the pavement rehabilitation due to loss of service life. The equation was derived from basic engineering economics formulas; it has been shown to produce a reliable pay factor relationship provided the input values are reasonably accurate (*5*). The appropriate pay factor was calculated to be 0.63 using the following data:

$$\text{PF} = \frac{M_c + (R^{DL} - R^{EL}) * \left(R_1 + \dfrac{R_2 R^{EO}}{1 - R^{EO}}\right)}{M_c}$$

where

M_c = cost of hot mix = \$45/T,
R_1 = cost of first resurfacing = \$60/T,
R_2 = cost of second resurfacing = \$60/T,
D_L = design life = 10 years
E_L = expected life = 8 years (20 percent loss of service life due to poor compaction),
E_O = expected life of overlays = 10 years (single lift),
I_F = inflation rate = 3.0 percent,
I_N = interest rate = 7.0 percent, and
R = 1.03 / 1.07 = 0.96.

(Resurfacing costs include the cost of removal.)

Using this model, appropriate pay factor curves were plotted for different standard deviations measured from the 1992 data for pavement compaction. A computer simulation was then used to develop a continuous pay schedule. The expected payment curve shown was generated by computer program (*5,6*). For comparison, the payment equation has been plotted with the appropriate pay factor curves (Figure 2) and the expected payment curve (Figure 3).

Several key observations can be made from these curves:

1. The appropriate pay factors determined using the different standard deviation values show the relationships between uniformity (or variability) and the estimated PWL. It is apparent that as the lot standard deviation increases, the price adjustment increases (pay factor decreases).
2. The payment curve shows that the minimum pay factor was determined to be 0.65 based on an expected life–to–design life ratio of 0.80. A bonus of 3 percent will be paid for lots exceeding the desired compaction level. A lot is deemed to be rejectable and may be subject to repair if the PWL is less than 50.
3. A comparison of payment curve and the expected pay factor curve revealed that at acceptable quality level (AQL), the actual payment is artificially higher than the expected payment curve. This was done to eliminate any bias by imposing price adjustments for material considered to be of AQL or better (i.e., 90 PWL). The flat area on the pavement compaction curve between 90 and 95 PWL was created primarily to simplify administration. This area would allow for a zone in which the material is accepted at full price. The difference in this area of the curve is small for an expected payment of 95 PWL. The remainder of the payment schedule curve matches the expected payment curve very closely up to about 70 percent PWL (30 percent defective), after which the payment curve separates from the expected payment curve. This separation is attributed primarily to the number of samples tested (*n*) and decreases as *n* becomes larger. The noted difference can be justified by the highway agency to account for the future maintenance, engineering, and administrative cost associated with the acceptance of deficient material at a reduced price (*5*).

ACCEPTANCE PROCEDURE

Acceptance for pavement compaction is to be based on the estimated PWL in accordance with the upper and lower specification limits provided in Table 2. The limits were determined on the basis of the data analyzed and reflect the level of compaction attained for the various mix types.

The lot mean and standard deviation will be used to estimate the lot PWL. The PWL will be calculated by determining the quality indexes, Q_l and Q_u, based on the following equations:

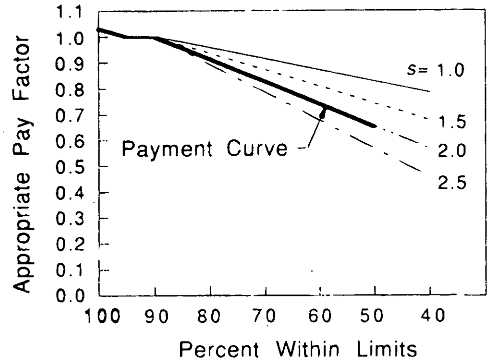


FIGURE 2 Appropriate pay curves.

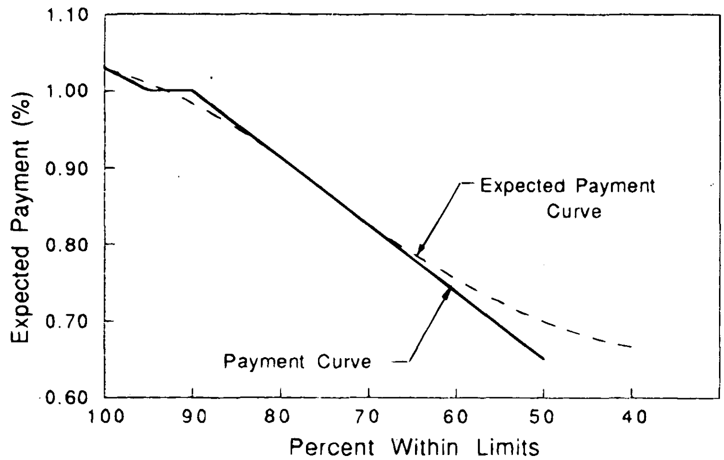


FIGURE 3 Expected payment curves.

$$Q_l = \frac{\overline{X} - \text{LL}}{s} \tag{1}$$

$$Q_u = \frac{\text{UL} - \overline{X}}{s} \tag{2}$$

where

Q_u = upper quality index,
Q_l = lower quality index,
UL = upper limit,
LL = lower limit,
X = lot mean, and
s = lot standard deviation.

The quality indexes are then used to determine the percentage of material above the lower limit and the percentage of material below the upper limit from the quality index table in the special provision. The total PWL for the lot is calculated using Equation 3:

$$\text{PWL} = (P_L + P_U) - 100 \tag{3}$$

where P_L is percentage within lower limit and P_U is percentage within upper limit.

SIMULATION ANALYSIS

To analyze the impact of the proposed specification, simulations were carried out using the 1992 compaction data. Although these simulations provide a good approximation of the impact of the proposed specification, it should be understood that three factors may slightly misrepresent the outcome of simulations. First, the previous compaction specification in place at the time that the data were collected did not effectively consider the uniformity of the mix. Second, the old system did not deter the contractor from overcompacting the mat. Third, the lack of a clear procedure to account for poor compaction did not encourage the contractor to provide good process control.

The compaction results from 11 DFC contracts were analyzed to stimulate the effect of the specification. A summary of the results from this simulation is presented in Figure 4. The results indicate that 22 percent of the lots would receive a bonus and 65 percent of the lots would receive less than full payment (before retesting or repairs). In all cases, these lots failed to comply with the specification requirements because of low compaction. The average compaction payment factor for a lot would be about 91.5 percent of the contract price. From the 1992 data, the overall price adjustment for DFC is expected to be \$3.40/T. The high number of price-reduced and rejectable lots is a concern to both the ministry and the hot-mix industry. Unfortunately, the severe consequences of accepting hot mix with less than 89.5 percent compaction (greater than 10.5 percent air voids) restrict the ministry's ability to lower the specification limits.

The compaction results from 21 HL-4 contracts were analyzed to simulate the effect of the specification. A summary of the results from this simulation is presented in Figure 5. Sixty-four lots (42 percent) would receive a bonus, and 68 lots (44 percent) would receive less than full payment. Of the 12 rejectable lots, 9 were determined to be rejectable primarily because of overcompaction, which was not addressed by the previous specification. Overall, the results indicate reasonable compliance, with more than half the lots being accepted at full price or receiving a bonus. The average compaction payment factor for a lot would be about 95.9 percent of the contract price. The estimated overall price adjustment for HL-4 is anticipated to be approximately \$1.50/T.

TABLE 2 Tolerance Limits for Pavement Compaction

Mix Type	LL(%)	UL(%)
HL1, HL2, HL3, HL3A, HL4, HL8, MDBC, RHM, hot in-place recycled mix and hot in-place recycled mix with integral overlay	91.5	97.0
HDBC	90.5	97.0
DFC	89.5	98.0

OC CURVES

A computer program was used to generate 12,000 independent random compaction results based on the various population character-

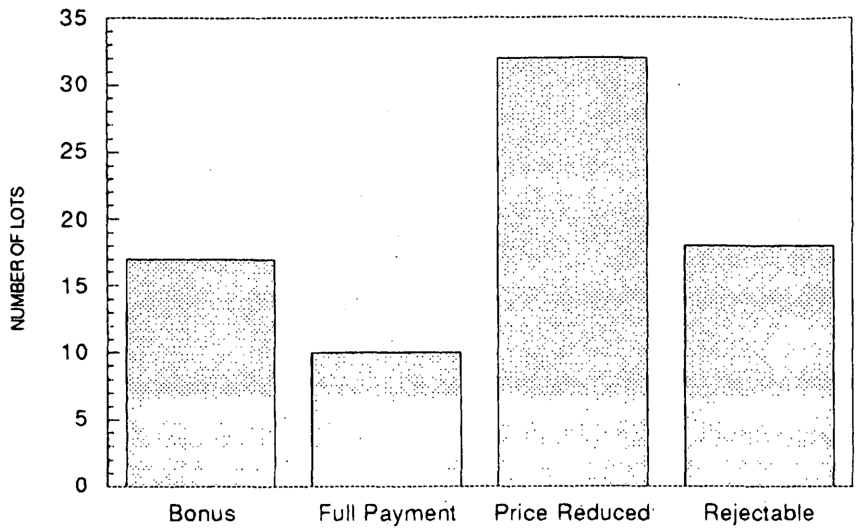


FIGURE 4 Simulation results for DFC mixes.

istics for each of the mix classifications. The results were separated into 2,000 lots with six samples each to form points for the OC curves shown in Figure 6.

The OC curves indicate that the acceptance plan worked as intended. The expected pay factors are high for population means at the desired level of compaction with low variability (standard deviations). Accordingly, the expected pay decreases as the variability increases and the population mean deviates from the expected target. The specification should provide an incentive for contractors to reduce variability and achieve an overall better end product.

SUMMARY

The primary objective of this paper was to develop ERSs for pavement compaction. The acceptance plan has been described in detail. The new specification is based on a PWL concept that can be adopted for most materials used in highway construction.

The data compiled from the 1992 contracts were used to determine the acceptance limits and to establish a continuous price adjustment schedule. On the basis of these data, the overall price adjustment for conventional mixes was estimated to be on the order of \$1.50/T.

To verify that the plan can be implemented and is fair to both MTO and the road builders, it was agreed to run a field simulation in fall 1993, incorporating different mix types and paving conditions across the province. The simulation would give MTO construction personnel and contractors a chance to gain experience with the acceptance plan and, more important, to determine if any modifications are required to the specifications. Modifications could entail loosening or tightening the acceptance limits, changing the sample size, or reducing or increasing payment for a given PWL.

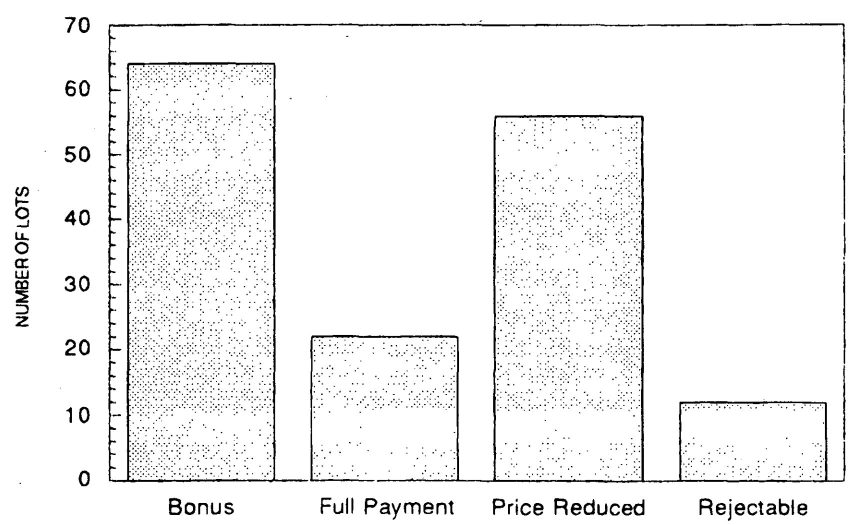
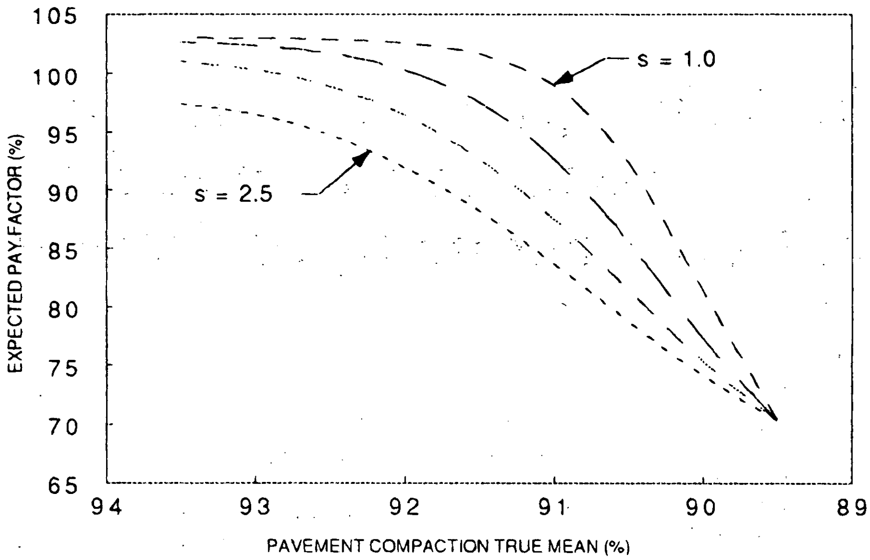


FIGURE 5 Simulation results for HL-4 mixes.



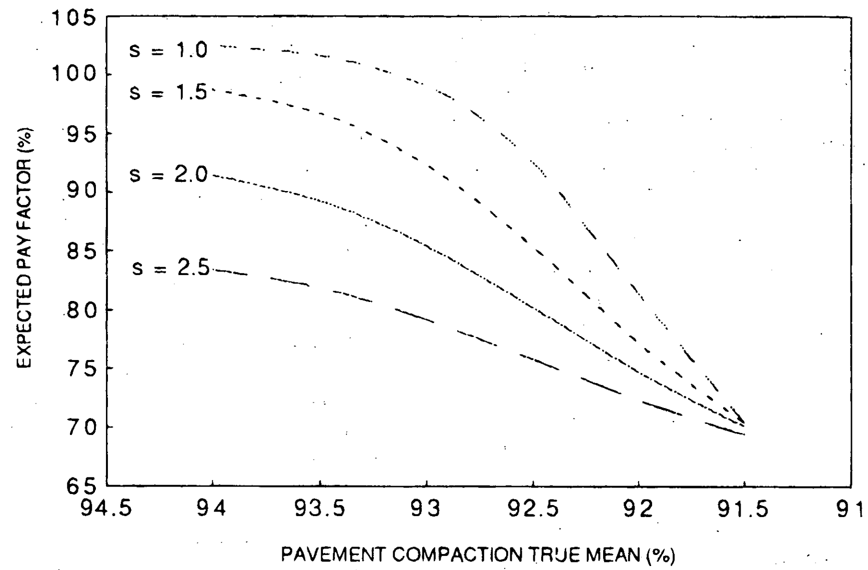


FIGURE 6 OC curves for pavement compaction: *top,* DFC; *bottom,* all other mixes.

RECOMMENDATIONS

The 1993 field simulation results should be reviewed, and if the specification shows satisfactory performance, it should be implemented with full price adjustment on selected contracts advertised in the 1994 construction season. The proposed ERS should be applied to all mix types analyzed in this study.

REFERENCES

1. Burati, J. L., Jr., and S. N. Amirkhanian. *Review of Ministry End Result Specifications.* Report MI-140. Engineering Materials Office, Ontario Ministry of Transportation, 1990.
2. *Hot-Mix Asphalt Paving Handbook.* Publication UN-13 (CEMP-ET). U.S. Army Corps of Engineers, 1991.
3. Linden, R. N., J. P. Mahoney, and N. C. Jackson. Effect of Compaction on Asphalt Concrete Performance. In *Transportation Research Record 1217,* TRB, National Research Council, Washington, D.C., 1989, pp. 20–28.
4. 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.
5. Weed, R. M. *Statistical Specification Development.* FHWA, U.S. Department of Transportation, 1982.
6. Weed, R. M. *OCPLOT PC Program to Generate Operating Characteristic Curves for Statistical Construction Specifications.* Draft Report. New Jersey Department of Transportation, Trenton, June 1994.

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

OCPLOT: PC Program To Generate Operating Characteristic Curves for Statistical Construction Specifications

RICHARD M. WEED

The performance of the nation's highway system is inexorably linked to the quality of design and the quality of construction. To control the quality of construction, transportation agencies have developed elaborate quality-assurance programs, most of which employ end-result specifications that rely on statistical sampling and acceptance procedures to ensure that the work is done in accordance with the plans and specifications. Whether the acceptance procedure leads to a simple pass-or-fail decision or an adjustment in contract price, the proper design of such plans is critical to their performance. Poorly conceived plans may be either totally ineffective or impractically severe, and both extremes have been found in published and proposed national standards. To encourage the proper design of plans that are both effective and fair, an interactive PC program has been developed that enables the user to construct operating characteristic curves to analyze the performance of a wide range of acceptance plans. An example is presented to demonstrate the versatility of the program and the ease with which it can be applied.

One of the nation's most valuable assets is the network of roads and bridges linking the suppliers of goods and services with their customers. State transportation agencies, which bear most of the responsibility for maintaining the highway system in good working order, have responded by developing elaborate programs to ensure that adequate quality is achieved and maintained. Most agencies rely on end-result specifications that use statistical sampling and acceptance procedures to make sure that the work is done in accordance with the plans and specifications. The acceptance tests are performed on random samples taken at either the job site or the supplier's plant. The acceptance procedure may lead to a simple pass-or-fail decision or it may lead to an adjustment of contract price. Whichever method is used, the proper design of such plans is critical to their performance. Poorly conceived plans may be totally ineffective or impractically severe, and both extremes have been found in published national standards (*1*).

OPERATING CHARACTERISTIC CURVES

Although it is not yet used widely in the highway field, there exists a well-established analytical procedure to check that an acceptance procedure will be both fair and effective. The procedure consists of constructing the operating characteristic (OC) curve (*1*, Part 3, Item 6), a graphical representation of the discriminating power of the acceptance procedure.

New Jersey Department of Transportation, 1035 Parkway Avenue, CN 600, Trenton, N.J. 08625.

Even though the acceptance procedure or pay equation spells out precisely the decision to be made for any level of measured quality, there is always some degree of uncertainty in the quality measurement itself. This uncertainty occurs because only a small fraction of each lot is sampled and tested and the test procedures themselves are not perfectly repeatable. The OC curve, if constructed properly, is capable of accounting for this uncertainty.

A conventional OC curve is shown in Figure 1. The probability of acceptance is indicated on the *y*-axis for the range of quality levels indicated schematically on the *x*-axis. The contractor's risks of having good—acceptable quality level (AQL)—work rejected and the agency's risk of accepting poor—rejectable quality level (RQL)—work are both illustrated in this figure.

Figure 2 presents an OC curve constructed for a statistical specification with an adjusted pay schedule. Quality levels are indicated on the *x*-axis in the usual manner but, instead of probability of acceptance, the *y*-axis gives the expected pay factor.

Although the risks have a slightly different interpretation in Figure 2, essentially the same information is provided. In this example, AQL work receives an expected pay factor of 100 percent, as desired, whereas RQL work receives an expected pay factor of 70 percent. Presumably, the specification is based on a quantitative performance model (*2*) that has enabled the highway agency to estimate the amount of payment to be withheld to cover the anticipated cost of future repairs (*1*, Part 3, Item 10). It can also be seen in this figure that truly superior quality may receive a bonus pay factor up to 102 percent.

The opportunity to earn at least some degree of bonus payment is necessary in order for a statistical acceptance procedure to pay an average of 100 percent when the work is exactly at the AQL. Because of the inherent variability of any sampling and testing process, some samples will underestimate the quality while others will overestimate it. Unless the acceptance procedure is designed to allow bonuses and reductions to balance out in a natural way, the average pay factor will be biased downward at the AQL and acceptable work may be penalized unfairly. The failure to award an average pay factor of 100 percent at the AQL, even by only 1 or 2 percent, can result in many thousands of dollars of unwarranted pay reductions throughout the course of a construction season.

PROBLEMS WITH OVERLOOKING OC CURVES

The following two examples are taken from national standards before their recent correction. They illustrate the two extremes—unduly lenient and unduly severe—that can occur when acceptance procedures are based on faulty premises and are not subsequently checked by constructing the OC curves.

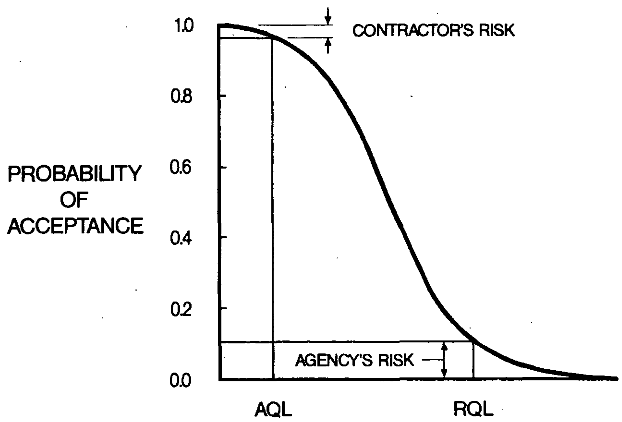


FIGURE 1 Conventional OC curve.

The first example is taken from a generic acceptance procedure that contains a number of desirable features. It uses the statistical measure of percent within limits (PWL) to account for both mean level and variability. [Percent defective (PD) is equally suitable]. The PWL estimate is computed by the standard deviation method which, because it is more statistically efficient, requires smaller sample sizes than plans based on the range (R). The procedure also includes a bonus provision, an essential feature if plans of this type are to operate fairly. And although it does not use a pay equation, which avoids potential disputes over test precision because of the smooth progression of payment as the quality varies, this procedure is nearly as effective because it uses a pay schedule with many small steps.

Despite these advantages, however, this procedure had one major shortcoming: it paid an average of nearly 104 percent for work that just met the AQL. In fact, it was so lenient that it paid an average of 100 percent for quality that was substantially below the AQL. Table 1 presents the OC curve for this plan in tabular form for a typical sample size of $N = 5$.

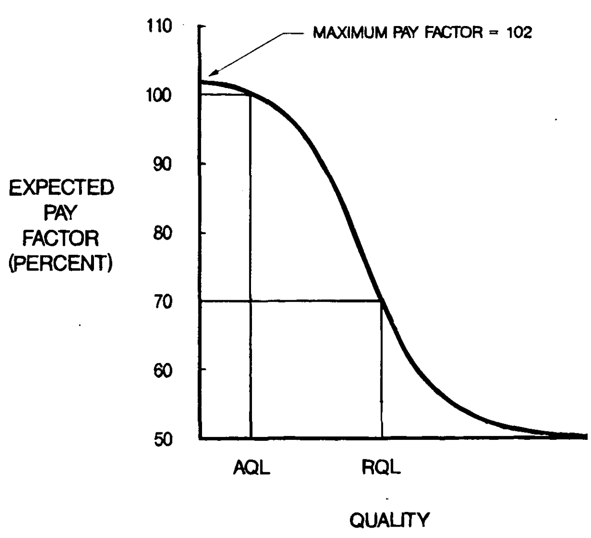


FIGURE 2 Typical OC curve for statistical acceptance procedure with adjusted pay schedule.

TABLE 1 OC Values of Lenient Acceptance Plan

PERCENT WITHIN LIMITS (PWL)	AVERAGE PAY FACTOR (PERCENT)
100	105.0
95 (AQL)	103.7
90	102.2
85	100.5
80	97.8
75	94.3
70	89.9

It is unlikely that any agency would want to use an acceptance plan that provides this degree of overpayment at the AQL. This example illustrates in a dramatic way the value of constructing the OC curve as part of the specification development process so that problems of this nature can be detected and corrected before implementation.

Although it is not known that this was the case, it is possible that this problem was the result of a common misconception about risk analysis as it applies to adjusted payment procedures. For pass-or-fail acceptance plans that produce OC curves of the type shown in Figure 1, the contractor's risk may typically be about 0.01 to 0.05. In other words, there is a 0.01 to 0.05 probability that work that is truly acceptable will be rejected. If the developer of a pay adjustment plan with a bonus provision were to attempt to control the risk of obtaining a pay factor of less than 100 percent at the AQL at a similarly low level, the vast majority of pay factors for acceptable work would exceed 100 percent and the overall average pay factor would be well above 100 percent, as happened in this example. For the acceptance procedure to perform properly, the risk of a pay reduction at the AQL must be approximately 0.50 so that, over an extended period of time, the pay factors for AQL work are split about evenly between bonuses and reductions.

The second problematical example demonstrates how unduly severe an acceptance plan can be if it is not designed properly. This example also includes most of the desirable features in that it uses the standard deviation method to estimate the PWL and a pay equation to compute the lot pay adjustment. However, the maximum pay factor was limited at 100 percent and, because this eliminated the opportunity to receive bonus payments, the procedure was not capable of paying an overall average of 100 percent when the work was precisely at the AQL.

Equation 1 is the basic form of the pay equation that was used. Because the specified AQL for this example is a PWL value of 90, this pay equation can also be expressed in the forms given by Equations 2 and 3, either of which is suitable for analysis with program OCPLOT.

$$\text{PAY REDUCTION} = \text{PWL}_{\text{specified}} - \text{PWL}_{\text{computed}} \qquad (1)$$

$$\text{PF} = 10 + \text{PWL} \qquad (\text{maximum} = 100) \qquad (2)$$

$$\text{PF} = 110 - \text{PD} \qquad (\text{maximum} = 100) \qquad (3)$$

where PF equals the pay factor.

The actual input and output stages with program OCPLOT will be shown later. For now, just the tabular form of the OC curve is presented in Table 2, computed for a typical sample size of $N = 5$.

TABLE 2 OC Values of Severe Acceptance Plan

PERCENT WITHIN LIMITS (PWL)	AVERAGE PAY FACTOR (PERCENT)
100	100.0
90 (AQL)	95.4
80	87.8
70	78.4
60	68.9
50	59.9

Table 2 indicates that a contractor who performs consistently at the AQL will receive an average pay reduction of nearly 5 percent. To emphasize the impact that this would have on the construction industry, this means that a contractor responsible for $10 million worth of pay adjustment work over the course of a construction season would be penalized $500,000 for successfully providing the level of quality that was defined as acceptable in the contract documents. This obviously is misleading and unfair.

Two simple steps will correct this problem. The first is to include a bonus provision as part of the acceptance procedure. In Equations 2 and 3, for example, this would mean removing the limitation that the maximum pay factor cannot exceed 100 percent. The magnitude of the maximum pay factor and the slope of the pay equation should be consistent with established (or estimated) performance relationships and the anticipated economic consequences of any departures—increases or decreases—from the specified AQL. The second step is to construct the OC curve to make sure that the resultant acceptance plan is neither too lenient nor too severe, as was the case for these two examples.

FEATURES AND CAPABILITIES OF OCPLOT

The type of acceptance plan represented by Equations 1–3 is one of literally hundreds that are capable of analysis with OCPLOT. Figure 3 lists some of the options that may be selected and the versatility of the program is apparent from the many ways in which these selections might be combined.

The programming is done in Microsoft QuickBASIC. It is highly structured and modular—consisting of three primary analytical modules, four auxiliary modules, and more than two dozen subroutines—and requires somewhat less than 1 megabyte of disk space. The program and its support modules may be loaded onto the hard drive or run from a diskette from the drive in which the diskette is placed. When the name OCPLOT is entered, preliminary screens identify the program as part of FHWA Demonstration Project 89 on Quality Management (*3*) and provide basic operational information.

ACCEPTANCE METHOD
- Pass/Fail ------------------------ TYPE OF PLAN
  - Attributes
  - Variables
- Pay Adjustment

QUALITY MEASURE
- Percent Defective (PD)
- Percent Within Limits (PWL)

LIMIT TYPE
- Single-Sided
- Double-Sided

PAY EQUATION TYPE
- Linear/Nonlinear ------------------------ Enter Values

MAXIMUM PAY FACTOR
- Yes ------------------------ Enter Value
- No

ACCEPTABLE QUALITY LEVEL (AQL)
- Enter Value

REJECTABLE QUALITY LEVEL (RQL)
- Enter Value

RQL PROVISION
- Yes ------------------------ Enter RQL Pay Factor
- No

RETEST PROVISION
- Yes ------------------------ INITIAL TESTS
  - Combined
  - Discarded
- No

SAMPLE SIZE
- Enter Value(s)

FIGURE 3 Available selections in OCPLOT.

Once this is complete, the first menu appears on the screen. Figure 4 shows this menu as it appears after all the selections have been made to analyze the acceptance procedure in the form represented by Equation 2. The various items appear on the menu one at a time in a logical sequence, and later items are dependent on the responses to earlier ones. For example, if a pass-or-fail type of acceptance method had been selected in response to the first query, a different set of subsequent queries would have followed. Besides the many combinations of features that can be accommodated, there is considerable latitude in selecting the values of specific parameters for any particular acceptance plan.

To the extent possible, an attempt has been made to include various checks in the programming to anticipate and avoid a variety of potential problems. For cases in which it is possible to know in advance that certain input values are improper, appropriate parts of the keyboard have been inactivated. In other cases, the program performs many internal checks to guard against the entry of inappropriate values. Depending on the degree of inappropriateness, two different responses may be displayed on the screen: a CAUTION message, color-coded yellow, that allows the user the option of either continuing or reentering a different value, and a WARNING message, color-coded red, that requires the user to enter a different value.

For example, the key with the minus sign is inactivated when the pay equation coefficients are selected and, when the input requires a choice among three menu items, only the keys representing the numerals 1–3 are active (except for <PrintScreen>, <ESC>, and <END>). If the user were to enter an RQL value that is unusually close to the AQL value, a yellow CAUTION message would appear and the user could enter either <ESC> to go back to select a different RQL value or any other key except <END> or <PrintScreen> to continue with the current selection. If the user

```
                    ENTER THE FOLLOWING INFORMATION

ACCEPTANCE METHOD                       ACCEPTABLE QUALITY LEVEL
   Pay Adjustment                          PWL = 90

QUALITY MEASURE                         REJECTABLE QUALITY LEVEL
   Percent Within Limits                   PWL = 50

LIMIT TYPE                              RQL PROVISION
   Single-Sided                            None

PAY EQUATION                            RETEST PROVISION
   PF = 10 + 1 PWL                         None

MAXIMUM PAY FACTOR                      SAMPLE SIZE
   PF = 100                                5

                       Press any key to continue

<ESC> = Back                                          <END> = Exit
```

```
        SELECT LEVEL OF PRECISION

  (1)  Low -- Faster Execution

  (2)  Intermediate

  (3)  High -- Slower Execution

             SELECTION ■

<ESC> = Back         <END> = Exit
```

FIGURE 4 First (*top*) and second (*bottom*) menus for OCPLOT.

were to enter the same value for the AQL and the RQL, a red WARNING message would appear and, when any key other than <END> or <PrintScreen> is pressed, the cursor would move back for another RQL selection. If the pay equation coefficients were chosen such that the pay factor could become negative at any point, the program would run, but a CAUTION message would appear, stating that all negative pay factors will be set equal to 0.

Once the entries in the first menu are complete, the user may elect to use the <PrintScreen> key to obtain a copy of the input selections. Striking almost any other key will cause the second menu in Figure 4 to appear.

Because the program uses computer simulation to analyze whatever acceptance procedure is specified, it is very computationally intensive. Selection 1 provides the fastest execution, which is useful for exploratory work but may not be good enough for a final result. When this level is selected, 200 sample sets of the desired size are generated randomly from a normal population for each of several known levels of quality. This process is far more thorough and many times faster than testing the acceptance procedure with actual field data. Each sample set is evaluated in accordance with the acceptance plan specified in the primary menu in Figure 4, and the results are stored in memory. This function provides the data base with which the acceptance procedure is analyzed.

Selection 2 provides an intermediate level of precision for which 1,000 sample sets are generated at each quality level. This level of precision is usually satisfactory to report as a final result, producing points on the OC curve representing either probability of acceptance or expected pay factor that are typically accurate to within about 1 or 2 percent. If still better precision is required, Selection 3 will cause 5,000 sample sets to be generated at each quality level. This level of precision tends to produce a very smooth line when the OC curve is plotted.

Once the precision level is selected from the second menu, the computational process begins. For either low or intermediate precision, OCPLOT displays detailed information at the two key points at which risk levels are usually expressed—the AQL and RQL—as shown in Figures 5 and 6. This display serves two important purposes. For users less familiar with statistical estimation procedures and acceptance plans, the graphical displays at the AQL and RQL are both informative and educational. It may come as a surprise to some, for example, how widely distributed the quality estimates are, especially for small sample sizes. For users more familiar with statistical acceptance procedures, these displays provide assurance that the simulation process is working properly. The actual displays on a color monitor are color-coded to clearly distinguish acceptable and rejectable test results and the corresponding pay factors.

It can be seen in Figure 5, for example, why the absence of a bonus provision in the pay equation causes the average pay factor to be well below 100 percent at the AQL. The population from which these data were generated is precisely at the AQL of PWL = 90. For the sample size of $N = 5$ and analysis at an intermediate level of precision, the 1,000 estimates of lot quality range from a minimum of about PWL = 48 to the maximum possible value of PWL = 100. It is predicted theoretically, and can be demonstrated empirically, that the average of the PWL estimates in the upper histogram in Figure 5 will be very close to the true value of PWL = 90 because the PD/PWL measurement process is an unbiased statistical estimation procedure. In the lower histogram in this figure, the corresponding pay factors range from a minimum of about 58 percent to the maximum of 100 percent that is permitted with this acceptance plan. As a result, the average pay factor is only 95.4 percent, even though all the samples were drawn from a population that was exactly at the level of quality that was defined as acceptable.

It can be seen in Figure 6 that when the true quality level is at the RQL of PWL = 50, the PWL estimates for a sample size of $N = 5$ cover the complete range from 0 to 100 percent, with the majority falling between about 20 and 80 percent. The average pay factor at the RQL is 59.9 percent, which may be appropriate, depending on the degree of economic loss that the agency believes it incurs when RQL work is accepted.

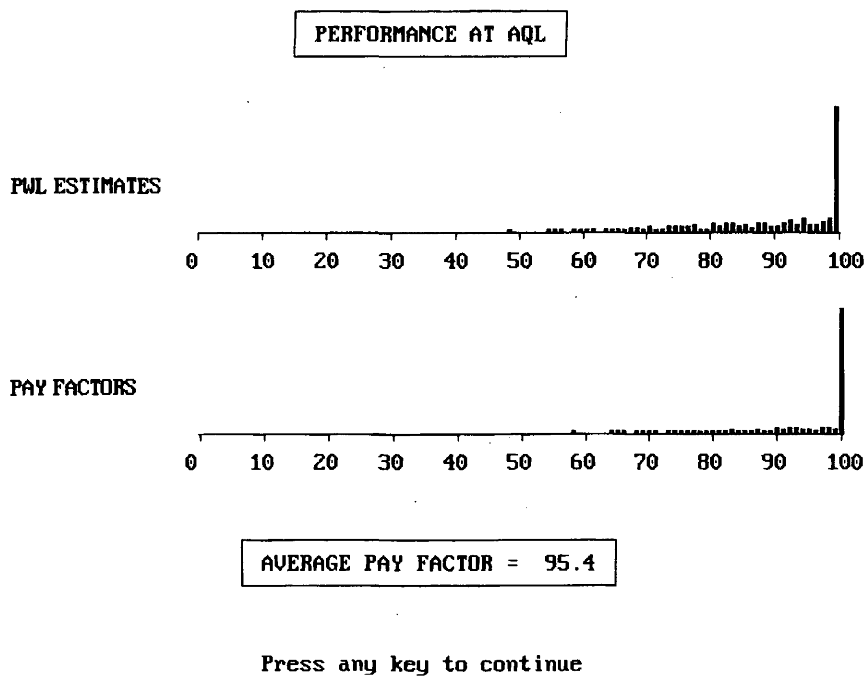


FIGURE 5 Display at AQL resulting from input shown in Figure 4.

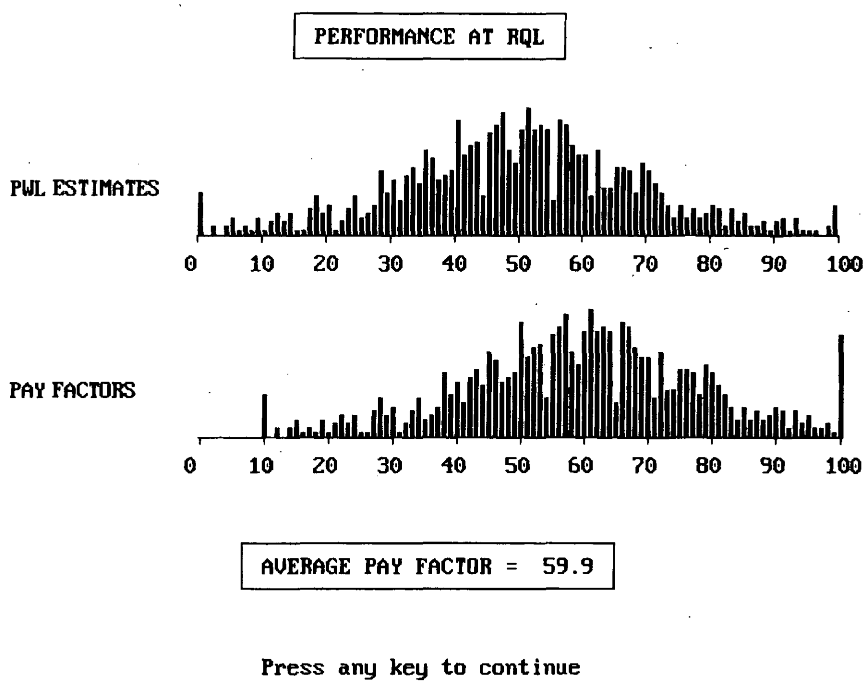


FIGURE 6 **Display at RQL resulting from input shown in Figure 4.**

Printouts of either Figure 5 or Figure 6 may be obtained provided the user's system has graphics capability, a commonly included feature with recent versions of DOS. A command similar to GRAPHICS [PRINTER TYPE] must be entered before running OCPLOT in order to obtain a printout using the <PrintScreen> key. A DOS manual should be consulted to obtain the appropriate syntax for the particular printer being used.

Although the AQL and the RQL are probably the most important points at which it is desired to know how the acceptance procedure will perform, it usually is useful to have a plot of the entire OC curve that provides a picture of the performance over the complete range of quality that might be encountered. The prompt at the bottom of the screen in Figure 6 instructs the user to strike any key to continue with this step, as shown in Figure 7. The *x-y* axes and the two previously calculated points at the AQL and RQL appear on the screen immediately. The remaining points appear one at a time at a speed determined by the level of precision that has been selected and the speed of the machine on which the program is being run.

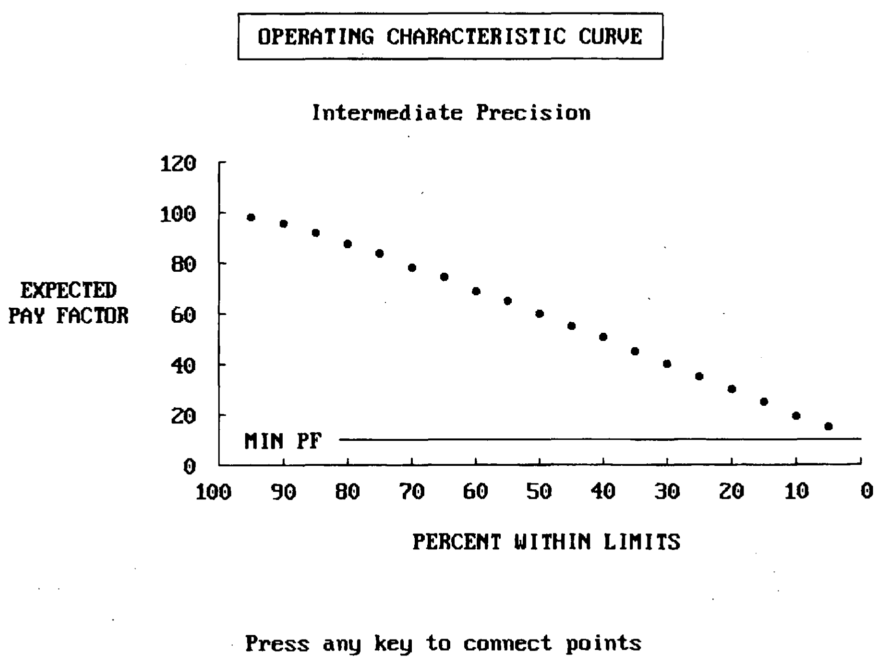


FIGURE 7 **Points on OC curve plotted by OCPLOT.**

For a 386 machine with a math coprocessor, this may require 1 or 2 min at low precision and 3 or 4 min at intermediate precision. With a 486 or faster machine, there is considerably less delay.

After all the points have been calculated and plotted, the user may strike any key to connect the points with a solid line. The next key stroke will add vertical and horizontal lines highlighting the performance of the acceptance plan at the AQL and RQL, as shown in Figure 8. And, like the histograms in Figures 5 and 6, any of these displays may be printed with the <PrintScreen> key, provided that graphics capability is present.

Following this display, striking a key will produce the menu shown in Figure 9. If the first item in this menu is selected, the output shown in Figure 10 is displayed. This feature permits the user to print out the values of the data points shown in Figure 7 from which the OC curve was constructed. The other selections in this menu make it possible to return to earlier points in the input stage of the program or to exit.

SOLUTION TO FAIRNESS PROBLEM

To demonstrate that the problem of paying less than 100 percent at the AQL can be corrected by allowing the pay equation to award bonus pay factors, another run was made with OCPLOT using the same input shown in Figure 4 except that no restriction was placed on the maximum pay factor.

Ordinarily, the maximum pay factor and the slope of the pay equation should be consistent with established (or estimated) per-

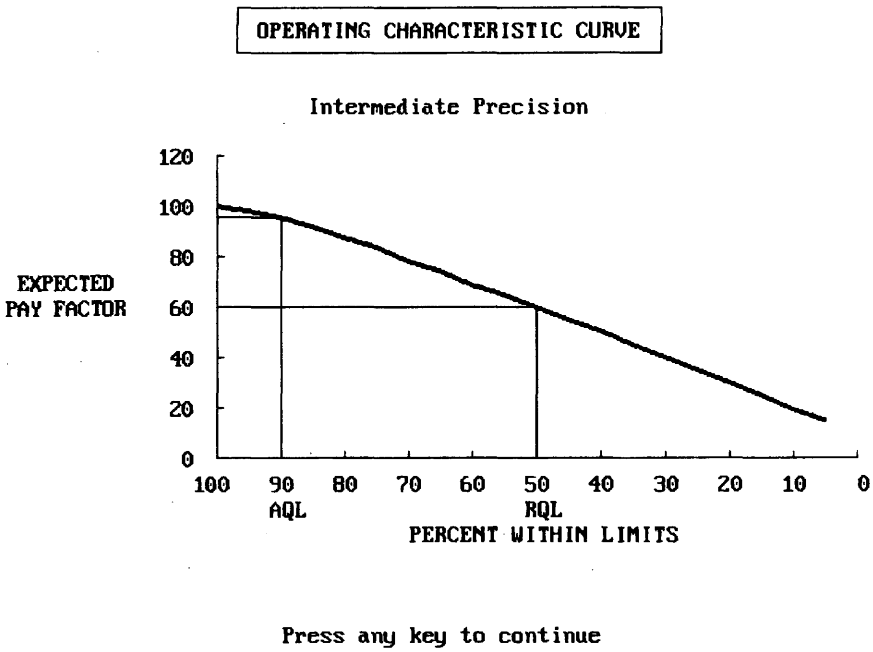


FIGURE 8 Display of OC curve with AQL and RQL performance highlighted.

SELECT DESIRED OPTION

(1) Display operating characteristic table

(2) Run again at different precision level

(3) Change some values and run again

(4) Run again with new input data

(5) Exit program

SELECTION ■

<ESC> = Back <END> = Exit

FIGURE 9 Third menu for OCPLOT.

PERCENT WITHIN LIMITS		EXPECTED PAY FACTOR
100.0		100.0
95.0		98.2
90.0	——— AQL ———	95.4
85.0		92.0
80.0		87.8
75.0		84.0
70.0		78.4
65.0		74.6
60.0		68.9
55.0		65.1
50.0	——— RQL ———	59.9
45.0		55.1
40.0		50.7
35.0		44.9
30.0		40.0
25.0		35.3
20.0		30.1
15.0		24.7
10.0		19.6
5.0		15.3

Press any key to continue

FIGURE 10 Display of numerical values of data points on OC curve.

formance relationships and the anticipated economic consequences of receiving quality levels other than the specified AQL. Because the pay equation used for this example is fairly steep, with a slope of 1.0, pay factors as large as 110 percent will be permitted when the restriction on maximum pay factor is removed. If a shallower slope had been used, a correspondingly lower maximum pay factor would have been appropriate.

The result at the AQL is shown in Figure 11. It can be seen that the PWL estimates are distributed almost exactly as they were in Figure 5 but that the pay factor estimates now range up to a maximum of 110 percent. Because of this, an average pay factor of almost exactly 100 percent has been achieved, as desired.

SUMMARY

An essential step in the writing of a statistical construction specification is the development of the OC curve. This is the only way to determine whether the acceptance procedure will distinguish properly between satisfactory and unsatisfactory work and award appropriate levels of payment. Two examples were presented to show that, in the absence of this step, the resulting acceptance plans could be either totally ineffective or impractically severe.

One reason that the construction of OC curves has not been a standard practice is that one of the most appealing measures of highway quality—PD, or its counterpart, PWL—is also one of the most complex to analyze. OCPLOT makes it possible for anyone with a minimal amount of statistical training to analyze a broad range of acceptance procedures of this type and, as such, provides a capability well beyond that previously available (*4*). For the less experienced user, the program provides additional guidance in the form of CAUTION and WARNING messages whenever a questionable entry is made.

This program, along with other quality-assurance software being distributed as part of FHWA Demonstration Project 89 (*3*), puts an enormous amount of analytical power in the hands of specification writers. It is hoped that the availability of this software will encourage a general upgrading of highway construction specifications, many of which may have shortcomings of the type illustrated in this paper, and that it will create a greater awareness of the need to develop acceptance plans that are both effective and fair.

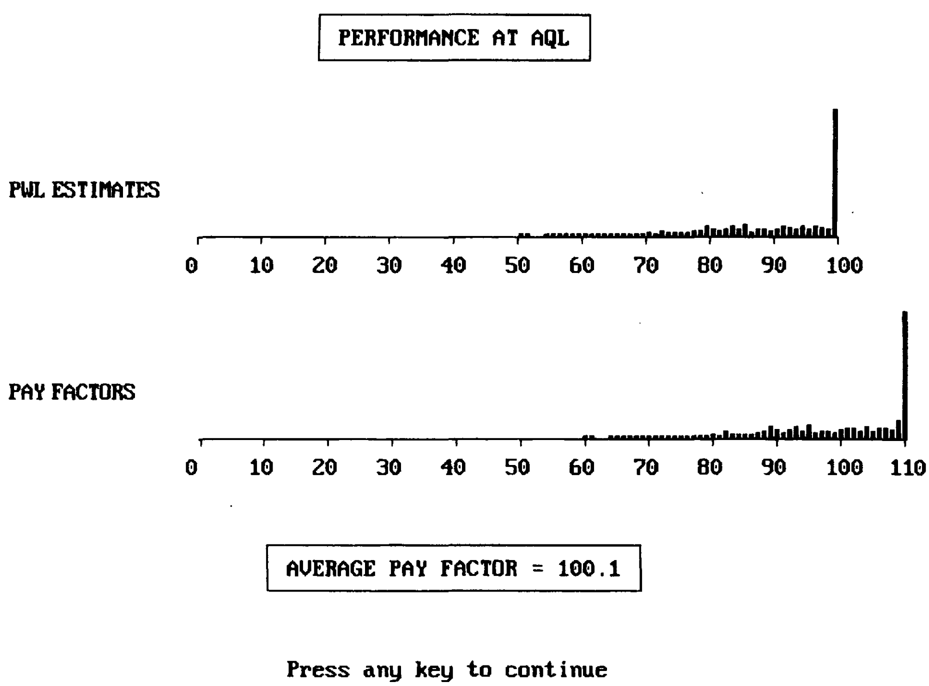


FIGURE 11 Demonstration that bonus provision produces average pay factor of 100 percent at AQL.

REFERENCES

1. 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.
2. *NCHRP Synthesis of Highway Practice 212: Performance-Related Specifications For Highway Construction and Rehabilitation.* TRB, National Research Council, Washington, D.C., 1995.
3. *Quality Management.* Demonstration Project 89. Demonstration Projects Division, FHWA, U.S. Department of Transportation, 1993.
4. Barros, R. T., R. M. Weed, and J. H. Willenbrock. Software Package for the Design and Analysis of Acceptance Procedures Based on Percent Defective. In *Transportation Research Record 924,* TRB, National Research Council, Washington, D.C., 1983, pp. 85–93.

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

Pavement Construction Smoothness Specifications in the United States

KHALED KSAIBATI, RICK STAIGLE, AND THOMAS M. ADKINS

In research conducted jointly by the University of Wyoming and the Wyoming Department of Transportation, current pavement construction smoothness specifications throughout the United States were collected and analyzed. A survey consisting of 13 questions dealing with pavement smoothness specifications, devices used for these specifications, and incentive and disincentive policies was sent out to all 50 state departments of transportation. Forty-five of the agencies responded to the survey. The responses were summarized in a computerized data base and analyzed for trends.

Road roughness is a major factor in evaluating the condition of highway pavement sections because of its effects on ride quality for road users and vehicle operating costs. In its broadest sense, road roughness has been defined as "the deviations of a surface from a true planer surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamics loads, and drainage" (*1*). Despite this broad description, the practice today is to limit the measurement of roughness qualities to those related to the longitudinal profile of the road surface that cause vibrations in vehicles using the road. Road roughness can also be defined as "the distortion of the road surface that imparts undesirable vertical accelerations and forces to the vehicle or to the riders and thus contributes to an undesirable, uneconomical, unsafe, or uncomfortable ride" (*2*). In general, road roughness can be caused by any of the following factors (*3*):

- Construction techniques that allow some variation from the design profile;
- Repeated loads, particularly in channelized areas, that cause pavement distortion by plastic deformation in one or more of the pavement components;
- Frost heave and volume changes due to shrinkage and swell of the subgrade; and
- Nonuniform initial compaction.

During the past three decades, several studies pointed out the major penalties of roughness to the user. In 1960 Carey and Irick showed that a driver's opinion of the quality of serviceability provided by a pavement surface is influenced primarily by roughness (*4*). Between 1971 and 1982, the World Bank supported several research activities in Brazil, Kenya, the Caribbean, and India, the main purpose of which was to investigate the relationship between road roughness and user costs. In 1980 Rizenbergs pointed to the following penalties associated with roughness: rider nonacceptance and discomfort, less safety, increased energy consumption, road-tire loading and damage, and vehicle deterioration (*5*).

Gillespie and Sayers examined the relationship between road roughness and vehicle ride to illustrate the mechanisms involved and to reveal those aspects of road roughness that play the major role in determining the public's perception of road serviceability (*6*). It is believed that the initial roughness of a pavement section will affect its long-term performance. Recently, a study conducted by Janoff suggested that initial pavement roughness measurements are highly correlated with roughness measurements made 8 to 10 years after construction.

Because of the importance of pavement roughness, most state highway agencies (SHAs) have established smoothness specifications for new pavement construction. Some SHAs require that a specific limit of smoothness be met, whereas others use a variable scale with price adjustment factors related to the degree of smoothness achieved. These price adjustments are based on the assumption that lower initial pavement roughness will result in better long-term pavement performance.

The University of Wyoming and the Wyoming Department of Transportation (DOT) are performing a joint research project to evaluate the effectiveness of smoothness specifications in Wyoming. As part of this evaluation, a nationwide survey was performed in spring 1994. The survey contained questions related to pavement smoothness specifications used by different SHAs. Most SHAs responding to the survey indicated their interest in learning about the findings of the survey. This paper summarizes the responses to the survey and shows the need for more uniform standards across the nation to accept pavement smoothness.

OBJECTIVES OF SURVEY

Copies of the smoothness specifications survey were mailed to all 50 SHAs in February 1994. The objectives of the survey were to

1. Identify the different roughness measurement devices used by SHAs to accept pavement smoothness for new construction,
2. Determine the acceptance limits for the various roughness measurement devices,
3. Identify SHAs that have incentive and disincentive policies for initial pavement smoothness,
4. Determine how SHAs developed their incentive and disincentive policies,
5. Estimate the percentage of pavement sections that qualified for incentives or disincentives in recent years, and
6. Evaluate the effectiveness of the various smoothness specifications.

K. Ksaibati and R. Staigle, Department Of Civil Engineering, University of Wyoming, P.O. Box 3295, University Station, Laramie, Wyo. 82071. T. M. Adkins, Wyoming Department of Transportation, P.O. Box 1708, Cheyenne, Wyo. 82002.

RESULTS FROM SURVEY

The construction smoothness survey included 13 questions aimed at satisfying the objectives stated. All states except California, Delaware, Missouri, Nevada, and Utah responded to the survey. The responses have been reduced and summarized in the sections that follow.

SHAs with Smoothness Specifications

Of the 45 SHAs that responded to the survey, only Massachusetts, Rhode Island, and Vermont indicated that they do not have any type of smoothness specifications. This rate implies that most highway agencies perceive initial pavement smoothness as being important.

Roughness Measurement Devices Used in Accepting Pavements

Many roughness measurement devices are on the market today. The accuracy and repeatability of measurements obtained with various devices vary from poor to excellent. A point of interest in this research project was to determine which measurement devices are being used for accepting new pavements. As indicated in Table 1, 30 out of 42 SHAs with smoothness specifications indicated that they use the California-type profilograph in accepting portland cement concrete (PCC) pavements. Five SHAs use the Rainhart profilograph, one uses the Mays meter, and four use other devices. The Michigan and Minnesota DOTs indicated that they use the GM profilometer (Michigan also uses the California-type profilograph). The New Jersey and Florida DOTs use a rolling straight edge for accepting concrete pavements. Alaska, Maine, and New Hampshire indicated that they do not build PCC pavements.

For the acceptance of new asphalt cement (AC) pavements, 15 SHAs indicated using the California-type profilograph and 16 use some form of a straight edge that varies in length between 3.05 and 7.62 m (10 and 25 ft). As presented in Table 1, five states used the Mays meter and four states use another type. Florida and New Jersey use rolling straight edges. Arizona uses the K.J. Law 690 DNC, and Michigan uses the GM profilometer and the California-type profilograph.

Minnesota, North Carolina, and Pennsylvania did not indicate the devices used for accepting asphalt sections. It should be mentioned that all SHAs using straight edges to accept asphalt pavements do not have any incentive and disincentive policies.

Acceptance Limits for Concrete Pavement

As presented in Table 2, most SHAs using the California profilograph specify a maximum smoothness acceptance limit of 110 or 158 mm/km (7 or 10 in./mi) for concrete pavement. An acceptance limit of 789 mm/km (50 in./mi) is used by Kansas DOT because of the elimination of the blanking band when reducing the pavement profile. Kansas does this to reduce the possibility of a long and low-amplitude wave being missed. Five states indicated using the Rainhart profilometer with various acceptance limits ranging from 63 to 189 mm/km (4 to 12 in./mi). Michigan and Minnesota indicated using the GM profilometer with acceptance limits of 49.8 ride quality index and 24 root mean square acceleration, respectively. Florida and New Jersey have an acceptance limit of 1 mm/m (1/8 in. in 10 ft) using a rolling straight edge. West Virginia uses the Mays ride meter with an acceptance limit of 1579 mm/km (100 in./mi).

Acceptance Limits for Asphalt Pavements

Most SHAs that pay pavement incentives or disincentives on asphalt pavements use the California-type profilograph. As indicated in Table 3, the consensus for an acceptance limit was 110 or 158 mm/km (7 or 10 in./mi). The rest of the SHAs indicated using a range of 47 to 189 mm/km (3 to 12 in./mi) except for Kansas, where a value of 631 mm/km (40 in./mi) is used for accepting

TABLE 1 Roughness Measurement Devices Used by SHAs To Accept Pavements

		Device Type				
		California Type Profilograph	Rainhart Profilograph	Straight Edge	Mays Meter	Others
Pavement Type	PCC*	AL, AZ, AR, CO, CT, HI, ID, IL, IN, IA, KS, LA, MD, MI, MS, MT, NE, NM, NY, ND, OH, OK, OR, PA, SD, TX, VA, WA, WI, WY	GA, KY, NC, SC, TN	0	WV	FL, MI, MN, NJ
	AC**	AL, ID, IL, IN, IA, KS, LA, MD, MI, NE, OH, OK, TX, VI, WI	0	AL, AR, CO, CT, HI, ME, MS, MT, NH, NM, NY, ND, OR, SD, WA, WY	GA, KY, SC, TN, WV	AZ, FL, MI, NJ

* PCC: Portland Cement Concrete

** AC: Asphalt Cement

TABLE 2 Acceptance Limits for PCC Pavements

	Acceptance Limits						
	63 mm/km 4 in/mile	79 mm/km 5 in/mile	95 mm/km 6 in/mile	110 mm/km 7 in/mile	158 mm/km 10 in/mile	189 mm/km 12 in/mile	789 mm/km 50 in/mile
California Type Profilograph	0	ID, ND	AL, CO, LA	AZ, AR, IA, MD, MS, NM, OH, OK, OR, WA, WY	HI, IL, MI, NE, PA, SD, TX, VA, WI	CT, IN, MT, NY	KS
Rainhart Profilograph	NC	0	0	GA	TN	KY, SC	0

TABLE 3 Acceptance Limits for AC Pavements, California Profilograph

	Acceptance Limits					
	47 mm/km 3 in/mile	95 mm/km 6 in/mile	110 mm/km 7 in/mile	158 mm/km 10 in/mile	189 mm/km 12 in/mile	631 mm/km 40 in/mile
California Type Profilograph	LA	AL, TX	ID, IA, MD, NE, OH, OK,	IL, MI, VA, WI	IN	KS

asphalt pavements with the same data reduction policy as described earlier for concrete pavements. Georgia, Kentucky, South Carolina, Tennessee, and West Virginia use the Mays meter with varying acceptance values, as indicated in Table 4. Four SHAs indicated using nonprofilograph devices for accepting asphalt pavements. Michigan uses both the California type profilograph and the GM profilometer; acceptance limits are 158 mm/km (10 in./mi) for the California profilograph and 49.8 ride quality index for the GM profilometer. Arizona has an acceptance limit of 1 mm/m (1/8 in. in 10 ft) using the K.J. Law 690 DNC. The rolling straight edge is used with acceptance limits of 2 mm (3/16 in.) for a 4.57-m (15-ft) length and 3 mm (1/8 in.) for a 3.05-m (10-ft) length by the Florida

TABLE 4 Acceptance Limits for AC Pavements, Mays Meter

Acceptance Limits			
Mays Meter Number			Rideability Index
35	40	100	3.6
GA, TN	SC	WV	KY

and New Jersey DOTs, respectively. A straight edge was the device of choice for all states without incentive and disincentive policies.

Incentive and Disincentive Policies

Incentive and disincentive policies used by different SHAs are of great interest to this research project. Table 5 gives the number of SHAs that have some sort of incentive or disincentive policy. Seventeen SHAs had incentive as well as disincentive policies for concrete pavements, but only 10 SHAs had both incentives and disincentives for asphalt pavements. Some SHAs had only incentive policies; others had only disincentive policies.

The information received on the actual incentive and disincentive policies varied greatly, with no more than two SHAs having similar policies. However, most SHAs had a similar upper-range adjustment price factor of 105 percent for incentives and a lower range of 90 percent for disincentives. Several SHAs would reduce the incentive percentage by 1 percent and increase disincentive percentages by 2 percent for every increase of 16 mm/km (1 in./mi). Examples of two incentive and disincentive policies are presented in Table 6. The immense variance of incentive and disincentive policies among SHAs indicates the variability of opinion on what profilograph index (PI) values indicate smooth or rough roads. More research is needed to determine the effect of PI values on the short- and long-term ridability of roads.

Development of Incentive and Disincentive Policies

As presented in Table 7, 22 SHAs indicated using engineering judgment in developing their current incentive and disincentive policies. Five SHAs based their specifications on research. However, states did not identify the type of research, length of study, or number of projects analyzed. Only three states indicated following AASHTO guidelines in the development of their specifications.

As indicated in Table 8, a majority of states indicated performing smoothness testing the same day or the day after pavement was laid for both asphalt and concrete pavements. Others responses indicated testing within 30 days or before the section is opened to traffic.

Performance and Percentages of Sections Qualifying for Incentive or Disincentive

Most SHAs said that they do not keep track of the percentages of pavement sections receiving incentives or disincentives. Those SHAs with good records showed significant differences in the percentages of sections receiving incentives or disincentives. The range of concrete sections that received incentives was 10 to 98 percent, whereas that incurring disincentives was 0 to 100 percent. New Jersey, the SHA reporting 100 percent disincentive on concrete pavements, requires less than 5 percent of the total lot to have surface variations greater than 1 mm/m (1/8 in. in 10 ft).

The variations among SHAs when considering asphalt pavements were as much as concrete pavements. The range of asphalt sections that received incentives was 15 to 95 percent; that incurring disincentives was 1 to 100 percent. Wisconsin, the SHA reporting 100 percent disincentive, assesses disincentives to any pavement that has a PI higher than 158 mm/km (10 in./mi) using a California-type profilograph.

Fourteen SHAs indicated observing roughness-related problems on sections that had received incentives. Some of these problems are due to the specifications, which do not always eliminate wheel chatter, or long wavelengths that create a roller coaster effect.

Effectiveness of Acceptance Specifications

The satisfaction of different SHAs with their current smoothness specifications was determined in this survey. Most SHAs rate their smoothness specifications as good or very good (Table 9). Only two states indicated poor satisfaction with their smoothness specifications.

CONCLUSIONS AND RECOMMENDATIONS

In this paper the responses of SHAs to a comprehensive pavement smoothness survey were summarized. These responses lead to the following conclusions:

- There is a great interest among SHAs in the subject of pavement smoothness specifications.

TABLE 5 SHAs with Incentive and Disincentive Policies

	Incentives and Disincentive	Incentives Only	Disincentive Only	None
PCC	AL, AZ, CT, IL, IA, KS, KY, MN, MT, NE, ND, OH, PA, SD, TX, VI, WI	MI, NM, OR, OK, WY	IN, LA, MD, MS, NJ, NY, SC, WV	AK, AR, CO, FL, GA, HI, ID, ME, MA, NH, NC, RI, TN, VT, WA
AC	AL, AR, IL, IA, KS, KY, NE, TN, TX, VA	MI, OK	IN, LA, MD, NJ SC, WI	AK, AR, CO, CT, FL, GA, HI, ID, ME, MA, MN, MI, MT, NH, NM, NY, NC, ND, OH, OR, PA, RI, SD, VT, WA, WV, WY

TABLE 6 Incentive and Disincentive Policies of Texas (*L*) and Alabama (*R*) DOTs

PI* mm/km (inches/mile)	Price Adjustment Factor
<47 (<3)	105%**
49 to 63 (3.1 to 4)	104%
65 to 79 (4.1 to 5)	103%
80 to 95 (5.1 to 6)	102%
96 to 110 (6.1 to 7)	101%
112 to 158 (7.1 to 10)	100%
159 to 174 (10.1 to 11)	98%
175 to 189 (11.1 to 12)	96%
191 to 205 (12.1 to 13)	94%
207 to 221 (13.1 to 14)	92%
223 to 237 (14.1 to 15)	90%
>237 (>15)	Correct

PI* mm/km (inches/mile)	Price Adjustment Factor
<47 (<3)	105%
49 to 95 (3.1 to 6)	100%
96 to 126 (6.1 to 8)	95%
128 to 158 (8.1 to 10)	90%
>158 (>10)	Correct

* Profilograph Index

** Percentage of Contract Unit Price

TABLE 7 Sources for Development of Incentive and Disincentive Policies

Research	AASHTO Guidelines	From Other States	Engineering Judgment Combined With One Other Category	No Specifications
CT, IL, OK, PA, WV	MS, OH, TX	MD, NY	AZ, HI, IN, IA, KS, KY, LA, MI, MN, MT, NE, NJ, NM, ND, OR, SC, SD, TN, TX, VA, WI, WY	AK, AR, CO, FL, GA, ID, MA, ME, NH, NC, RI, VA, VT

- Most SHAs use the California-type profilograph to accept pavement smoothness.
- A few SHAs still use response-type devices to accept pavement smoothness.
- The acceptance limits for pavement smoothness vary greatly among SHAs. Two sections with the same smoothness level may receive disincentives in one state and incentives in another.
- Most SHAs base their specifications on engineering judgment rather than research.
- Most SHAs are highly satisfied with their current smoothness specifications.

ACKNOWLEDGMENTS

This cooperative study was funded by the U.S. Department of Transportation University Transportation Program through the Mountain-Plains Consortium, the Wyoming Department of Transportation, and the University of Wyoming. The authors would like

TABLE 8 Timing of Pavement Smoothness Testing

		Time of Testing			
		Same Day	Next Day	End of Construction	Others
Pavement Type	PCC	AR, CO, OR, PA	AL, AR, CT, GA, ID, IL, IN, IA, KS, LA, MS, NM, NY, NC, ND, OH, OK, SD, TX, VA, WA, WI, WY	FL, KY, MN, NJ	HI, MD, MI, MT, NE, SC, TN, WV
	AC	AK, CO, KS, ME, NH, OR, VA, WA, WI	AL, AR, CT, GA, ID, IN, IA, LA, MS, NM, NY, ND, OH, OK, SD, TX, WY	FL, IL, KY, NJ	AZ, HI, MD, MI, MT, NE, SC, TN, WV

TABLE 9 Effectiveness of Smoothness Specifications as Rated by SHAs.

		Pavement Type	
		PCC	AC
Rating	Excellent	AR, KS, PA, VA	KS
	Very Good	AL, GA, ID, IA, LA, MN, MS, MT, NE, NY, OH, SD, TN, WV, WI	AL, AK, GA, ID, IA, KY, LA, MS, MT, NE, NY, OH, SD, TN, VA, WV,
	Good	AR, CT, FL, HI, IL, IN, MD, MI, NC, ND, OK, OR, SC, WY	AR, CT, FL, HI, IL, IN, MD, ME, MI, ND, OK, OR, SC, WI, WY
	Fair	KY, NJ	NJ
	Poor	CO, TX	CO, TX
	No Answer	NM, WA	AZ, NH, NM, WA,
	No Specifications	AL, ME, MA, NH, RI, VT	MA, MN, NC, PA, RI, VT

to express their appreciation to all the SHAs that responded to the survey.

REFERENCES

1. Sayers, M. W. *Development, Implementation, and Application of the Reference Quarter-Car Simulation.* ASTM, Philadelphia, A.A.
2. Hudson, W. R. Road Roughness: Its Elements and Measurements. In *Transportation Research Record 836,* TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
3. Yoder, E. J., and D. Hampton. *Pavement Profile and Roughness Measurements: A Review of Method.* Purdue University, West Lafayette, Ind., Jan. 1958.
4. Carey, W. N., and P. E. Irick. The Pavement Serviceability-Performance Concept. *Bulletin 250,* HRB, National Research Council, Washington, D.C., 1960, pp. 40–58.
5. Gillespie, T. D., and M. W. Sayers. Role of Road Roughness in Vehicle Ride. In *Transportation Research Record 836,* TRB, National Research Council, Washington, D.C., 1981, pp. 15–20.
6. Rizenbergs, R. L. Roughness of Pavements: An Overview. Presented to ASTM Committee E-17 on Traveled Surfaces Characteristics, Orlando, Fla., Dec. 1980.

The authors are solely responsible for the contents of this paper, and the views expressed do not necessarily reflect the views of the research sponsors.

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

Development of Air Voids Specification for Bituminous Concrete

RICHARD M. WEED

The New Jersey Department of Transportation (NJDOT) has been using statistical quality assurance (SQA) specifications for various construction items since the late 1960s. Throughout this period, there has been a continuing process leading to a better understanding of the operation and implementation of SQA procedures. The NJDOT specification for air voids in bituminous concrete was one of the first to be developed and, as such, was a prime candidate for upgrading. A major change is to base the acceptance procedure on the percentage defective rather than the average of the test values in order to control both the level and the variability of the air voids in a statistically efficient way. Doing this required new definitions of the acceptable and rejectable quality levels (AQL and RQL) and a reexamination of the adjusted pay schedule to be applied when other than AQL work is received. It was decided to use a positive incentive (bonus) provision for superior quality, an approach that has worked well with other recently developed NJDOT specifications. Another change is to use a continuous (equation-type) pay schedule to provide a smooth progression of payment as the quality varies, thus avoiding potential disputes over measurement precision when a quality estimate falls just on one side or the other of a boundary in a stepped pay schedule. The various developmental steps are described, including the construction of the operating characteristic curve to verify the performance of the specification and the field trials leading to its successful implementation.

The AASHO Road Test provided a wealth of statistical data in the early 1960s that could be used to relate various construction quality measures to performance. As did several other states at the time, New Jersey began to explore the use of this information to develop specifications that described the desired quality in statistical terms. This approach turned out to be effective, but it was discovered that for most construction items, it was not possible to define a single level of quality that clearly separated acceptable and unacceptable work. It was possible, however, to define a high level of quality that was clearly acceptable (AQL) and a substantially lower level that was clearly rejectable (RQL). In between these two extremes, work was judged to be sufficiently defective that it did not deserve full payment but not so defective that removal and replacement were warranted. Thus was born the concept of adjusted payment, which provided a convenient and practical way to accept minimally defective items for a prearranged level of reduced payment.

The New Jersey Department of Transportation (NJDOT) began to develop random sampling plans, statistical acceptance procedures, and adjusted pay schedules for various properties of bituminous concrete, including air voids, which are a surrogate measure for level of compaction. Because statistical quality assurance (SQA) was new to almost everyone in the transportation field, the early specifications were based on the simplest concepts. For example, the range was often used in favor of the standard deviation as a measure of variability because it was easier to understand and apply. Now it is recognized that a price in statistical efficiency must be paid for simplifications such as this, and that when acceptance procedures are based on the standard deviation, the same discriminating power can be obtained with a reduced sampling and testing effort.

The original air voids specification was simplified by using the process average as the acceptance parameter. The drawback of this is that it ignores variability. If the variability were to become unusually large, there could be a considerable amount of out-of-specification material even though the process average was at a normally satisfactory level.

Early SQA specifications typically assessed pay reductions for deficient quality but did not award extra payment for quality that was above that required. More recently, the use of positive incentive (bonus) payments for truly superior quality has been judged to be in the public interest (*1*); this is now a common practice in many states.

The first pay schedules to be developed typically had several distinct steps with a declining series of percentage pay factors corresponding to specific ranges of the quality measure. The problem with this approach is that whenever the true quality level falls close to one of the boundaries in such a pay schedule, the quality estimate may fall on either side of the boundary, primarily by chance. Depending on which side of the boundary the estimate falls, there may be a substantial difference in pay level, which can lead to disputes over test procedures, measurement precision, round-off rules, and so forth. Many recent SQA specifications avoid this problem by using continuous (equation-type) pay schedules that provide a smooth progression of payment as quality varies.

The development of the first SQA specifications was largely a trial-and-error process; several tries were often needed before a workable specification was obtained. Modern SQA specifications are the result of a continuing evolutionary process and contain many improvements and refinements not present in the earlier versions. As highway engineers have developed a better understanding of both the operation and implementation of SQA procedures, this newly acquired knowledge has been reflected in more effective acceptance procedures with properly balanced risks and fair and equitable adjusted payment provisions. As the level of sophistication has increased, the computer has emerged as a valuable aid in performing much of the developmental and analytical work. As the result of this steady progress, SQA specification writing is now far less of an empirical art and may begin to be regarded as a scientific process.

OBJECTIVES

One objective of an ongoing study was to establish a list of fundamental principles to guide the long-term upgrading of the NJDOT quality-assurance program. This goal has been completed, and a list

New Jersey Department of Transportation, 1035 Parkway Avenue, CN 600, Trenton, N.J. 08625.

TABLE 1 Effect of Compaction on Bituminous Pavement Life

AIR VOIDS (PERCENT)	REDUCTION IN PAVEMENT LIFE (PERCENT)		
	LITERATURE	SHA SURVEY	STUDY
7	0	7	0
8	10	13	2
9	20	21	6
10	30	27	17
11	40	38	--
12	50	46	36

of 28 basic SQA concepts has been published (*2*, Part 3). One of the tasks of the study described in this paper was to apply these principles to improve the NJDOT acceptance procedure for air voids in bituminous pavement.

It is widely recognized that level of compaction is one of the most important variables relating to long-term pavement performance. Table 1 has been reproduced from a recent publication (*3*) and summarizes the effect of compaction, measured in terms of air voids content, on the performance of bituminous concrete pavement.

The level of compaction can be controlled in either of two ways: by controlling the density of the pavement directly or by controlling the air voids content. NJDOT has elected to base its acceptance procedure on air voids because this approach accounts for additional important performance factors such as permeability, intrusion of road chemicals, oxidation, and potential for creating a hazardous condition by extruding asphalt onto the pavement surface when the voids content is too low.

Lot size is defined as either 4180 m^2 (5,000 yd^2) of bituminous concrete surface area of uniform thickness or 8360 m^2 (10,000 yd^2) of a pavement layer that is of variable thickness. For simplicity, the original acceptance procedure had been based on the average air voids content obtained from $N = 5$ cores, taken at random locations and evaluated in accordance with standard test methods. The average was required to be between 2.0 and 8.0 percent.

Provided that the five-sample average was between the limits of 2.0 and 8.0 percent, full payment was made. For averages falling outside these limits, the stepped pay schedule presented in Table 2 provided a decreasing level of payment. The minimum pay factor was PF = 80 percent, and there was no formal RQL provision requiring removal and replacement for extremely high voids content.

Although this procedure performed reasonably well initially, it was eventually discovered that there were an increasing number of cases in which the average air voids content was within the required range of 2.0 to 8.0 percent but that a substantial number of test values fell outside these limits, usually on the high side. Air voids measurements well above 10.0 percent were commonly encountered, and basic statistical reasoning suggests why this was the case. Provided that the average value is not unusually low, historical data have shown that percentage air voids is approximately normally distributed and the standard deviation may occasionally be as large as $\sigma = 2.0$ percent. If the average level were just within the upper limit of 8.0 percent, individual values could be as large as 14.0 percent, as shown in Figure 1. To emphasize the effect of this, an approximate smooth-curve fit of the information provided by Table 1 is drawn on this same figure. If Table 1 is correct, it is apparent that a substantial portion of the lot illustrated in Figure 1 would have a significantly shortened life.

TABLE 2 Original Pay Schedule for Percent Air Voids

AVERAGE AIR VOIDS (PERCENT)	PAY FACTOR (PERCENT)
0.0 - 1.4	80
1.5 - 1.9	90
2.0 - 8.0	100
8.1 - 9.0	95
9.1 - 10.0	90
Over 10.0	80

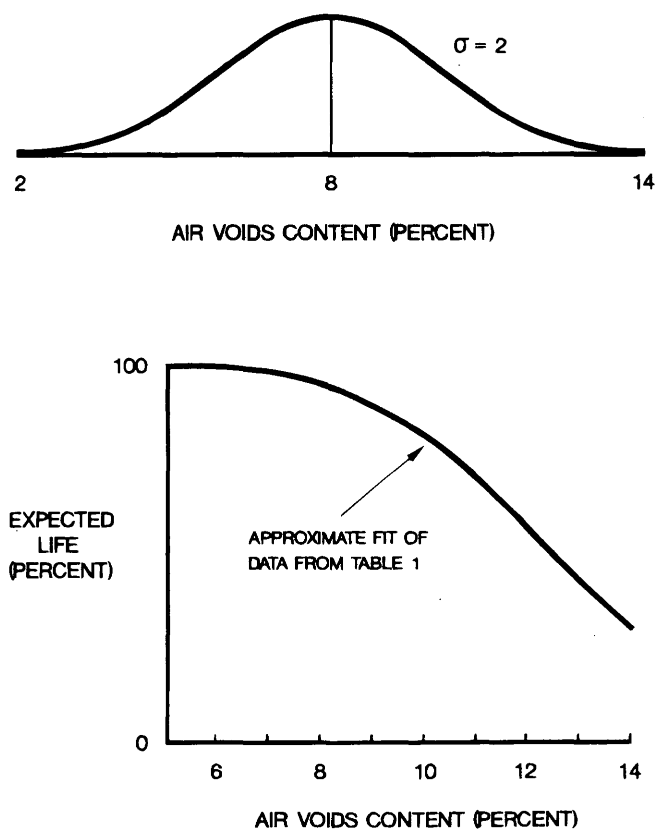


FIGURE 1 Air voids distribution with large standard deviation and associated life expectancy.

NEW TYPE OF SPECIFICATION NEEDED

It was concluded that a different type of acceptance procedure was needed, one that would provide an incentive to the contractor to control not only the average voids content but the variability as well. This need led to the decision to base the new specification on a different statistical measure of quality: percent defective (PD), the percentage of the lot falling outside specification limits.

It was decided to use the same specification limits, 2.0 and 8.0 percent air voids content, and define the AQL to be a lot for which no more than 10.0 percent of the material falls outside these limits. It was believed that this level of quality would ensure good performance and, at the same time, could be achieved by the construction industry. Another consideration was that this AQL had proven to be a practical choice for several other NJDOT specifications.

It was also necessary to define the RQL, the level at which the highway agency reserves the option to require removal and replacement, corrective action, or the assignment of a substantial pay reduction for the lot. As a general rule, RQL values must be set low enough that such drastic action is truly warranted. Since pavement failure does not pose a major safety hazard (such as the catastrophic failure of a bridge member), the RQL limit for percentage air voids has been set at a relatively high level of PD = 75. However, as an additional safeguard, another provision has been included that gives NJDOT the option of reevaluating the lot by means of additional cores whenever the estimated PD equals or exceeds 50.

DEVELOPMENT OF PAY SCHEDULE

In between the AQL of PD = 10 and the RQL of PD = 75, the work will be accepted at reduced payment. If specifications such as this are to be effective, the amount of payment reduction must be related in at least an approximate way to the economic loss expected to result from deficient quality. For truly superior levels of quality, ranging from PD = 0 to PD = 10, it is believed that additional service life of the pavement is obtained and, accordingly, an appropriate level of bonus payment is justified.

Besides providing an additional incentive to produce high quality, a bonus provision is an essential feature if specifications of this type are to perform fairly. Because there is inherent uncertainty in any sampling and testing process, some samples will underestimate the quality and others will overestimate it. Unless there is some way for bonuses and reductions to balance out naturally, the average pay factor will be biased downward at the AQL and acceptable work may be penalized unfairly. The importance of this feature becomes apparent when the operating characteristic (OC) curve (*2*, Part 3, Item 6) for the acceptance procedure is examined. A conceptual model of an OC curve for a statistical acceptance procedure with an adjusted pay schedule is shown in Figure 2.

For lower levels of quality (PD > 10), the same techniques that have proved successful in engineering economics and life-cycle cost analysis may be used to develop sound and defensible adjusted pay schedules. The fundamental assumption is that it is justifiable to withhold sufficient payment at the time of construction to cover the cost of future repairs made necessary by deficient quality. If a pavement is constructed with insufficient thickness, for example, it is possible to work backward through the design procedure to estimate the amount by which its service life will be shortened. The series of overlays that will commence at the end of the pavement's design life now must be initiated sooner. Since this will occur some

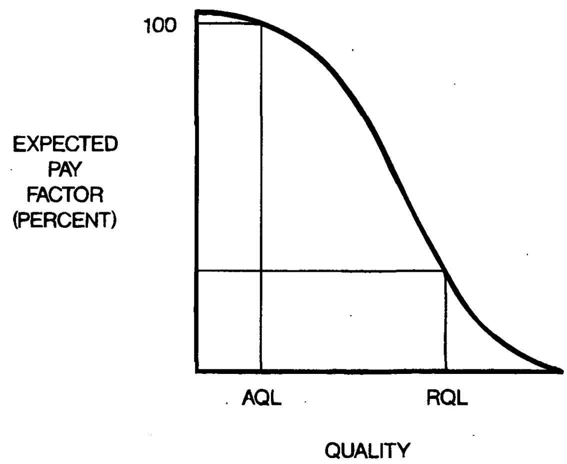


FIGURE 2 Conceptual model of OC curve for statistical acceptance procedure with adjusted pay schedule.

time in the future, both the interest rate on capital and the inflation of construction costs must be properly accounted for in order to estimate the monetary impact in terms of present worth. This approach is consistent with the legal principle of liquidated damages and is discussed in more detail in a recent publication (*2*, Part 3, Items 10 and 28), where Equation 1 is presented. Sensitivity tests have shown that pay schedules developed from this equation are relatively stable because costs as well as interest and inflation rates tend to track in parallel over long periods of time.

$$PF = 100\,[1 + (C_{overlay}/C_{pavement})\,(R^{L_{design}} - R^{L_{expected}})/(1 - R^{L_{overlay}})] \qquad (1)$$

where

- PF = appropriate percentage pay factor (dependent variable),
- $L_{expected}$ = expected life of pavement (years) (independent variable),
- $C_{pavement}$ = present unit cost of pavement (bid item only, same units as $C_{overlay}$),
- $C_{overlay}$ = present unit cost of future overlay (total in-place cost, same units as $C_{pavement}$),
- L_{design} = design life of pavement (years),
- $L_{overlay}$ = expected life of overlay (years),
- R = $(1 + R_{inflation}/100)/(1 + R_{interest}/100)$,
- $R_{inflation}$ = long-term annual inflation rate (%), and
- $R_{interest}$ = long-term annual interest rate (%).

To apply this procedure, it is necessary to determine appropriate values for unit cost, design life, and interest and inflation rates. It is also necessary to have a performance model that relates level of quality to expected service life in order to determine the value of $L_{expected}$ required by Equation 1.

Typical costs, design life, and expected overlay life were readily available from the design and construction units. Long-term interest and inflation rates could be estimated from our own records and published data. An approximate performance model, obtained by fitting a smooth curve through the data in Table 1, is shown in Figure 1 and is given by Equation 2:

$$PCTLIF = 100 - 0.756(VOIDS - 5)^{2.1} \qquad (2)$$

where PCTLIF is the life expectancy in percentage of design life, and VOIDS is the percentage air voids content.

Since this model curves downward at higher levels of percentage air voids, it reflects a greater degree of shortening of pavement life for the upper tail of any air voids distribution to which it is applied. As a result of this nonlinear effect, the use of a single average value for air voids content in Equation 2 will tend to overestimate the average life of the pavement to some degree. Tests with a numerical integration procedure suggest that the value of PCTLIF obtained in this manner may be overstated by about two or three units, depending on the mean and the standard deviation of the air voids distribution. However, this overstatement will not be accounted for in the analysis that follows; it will be treated only as further justification for a general tightening of the air voids acceptance procedure.

The procedure for developing the pay equation is as follows:

1. The appropriate percentage pay level is determined at two specific points: the AQL and RQL.
2. A trial pay equation is selected, including the minimum percentage pay level to be assigned when RQL work is allowed to remain in place. A linear pay equation usually will be sufficient.
3. The OC curve is constructed and the resulting average pay levels at the AQL and RQL are checked. This trial-and-error process is repeated until the desired results are obtained.

Like other NJDOT specifications, the AQL for the new air voids specification has been defined as a level of PD = 10. It is a fundamental requirement that when the work is precisely at the level defined as acceptable, the average pay factor must be 100 percent. Therefore, the location of the point at the AQL is already established.

To determine the appropriate pay level at the RQL of PD = 75, and for work of still lower quality that for practical reasons may be allowed to remain in place, the following values have been assumed:

$\sigma = 1.5$ (typical value for percentage air voids)
$C_{pavement} = \$6.91/m^2$ ($\$5.78/yd^2$) (assumes \$33/Mg bid price, average thickness of 8.9 cm, and compacted density of 23.5 kg/cm/m^2; in U.S. customary units, \$30/ton bid price, average thickness of 3.5 in, and compacted density of 110 lb/in/yd^2)
$C_{overlay}$ = total in-place cost of overlay of $\$11.96/m^2$ ($\$10.00/yd^2$)
L_{design} = 15 years
$L_{overlay}$ = 10 years
$R_{interest}$ = 8 percent (long-term annual rate)
$R_{inflation}$ = 4 percent (long-term annual rate)

Assuming a typical level of variability, the RQL of PD = 75 corresponds to an average air voids level of 9.0 percent. This value is substituted into Equation 2 to obtain an approximate value of PCTLIF = 86.1 percent which, when applied to the design life of 15 years, produces an expected life of 12.9 years. This value is then substituted into Equation 1 to obtain an appropriate pay factor for RQL work of about 74 percent. It is apparent that, on the basis of current costs and conditions, the previous minimum pay factor of 80 percent that was used with the original specification is not low enough to recoup the future anticipated costs associated with seriously defective work. In fact, it will be found that a pay factor of even lower than 74 percent must be assigned when RQL work is allowed to remain in place. Like the condition at the AQL where the opportunity to receive bonus pay factors greater than 100 percent allows the process to fairly award a long-term average pay factor of 100 percent, correspondingly lower lot pay factors are necessary at the RQL if the long-term average is to be at the desired level. It will be seen when the OC curve is constructed that an RQL pay factor of about 60 percent is required.

The third step of this process involves constructing the OC curve to determine how the pay equation will perform—that is, to verify that the desired average pay levels of 100 percent at the AQL of PD = 10 and 74 percent at the RQL of PD = 75 have been achieved. A computer program recently developed for FHWA Demonstration Project 89 on Quality Management (*4*) proved to be an extremely useful tool for this step; this program, OCPLOT, is described in another paper in this Record (see p. 18). The pay schedule given by Equation 3, when combined with an RQL pay factor of 60 percent, is shown in Figure 3 to produce a satisfactory OC curve. The actual wording of the specification and the necessary computations are described in the next section.

$$PF = 102 - 0.2\ PD \qquad (3)$$

NEW AIR VOIDS SPECIFICATION

It has been NJDOT practice to use somewhat relaxed pay schedules when phasing in new construction specifications in order to allow the construction industry time to become familiar with the new procedures. The pay equation included in this specification reduces all pay adjustments by exactly half. The pay reduction for RQL work that is allowed to remain in place is similarly about half of what the theoretical development suggests is warranted. It is planned to revert to more appropriate pay levels after sufficient experience has been gained with the new specification.

It is recognized that a slight improvement in statistical efficiency could be realized by using two different PD estimation tables: one for a sample size of $N = 5$ for initial tests, and another for $N = 10$ when the option to require a retest is exercised. The decision to use a single table is a concession to practicality that sacrifices very little in the way of performance.

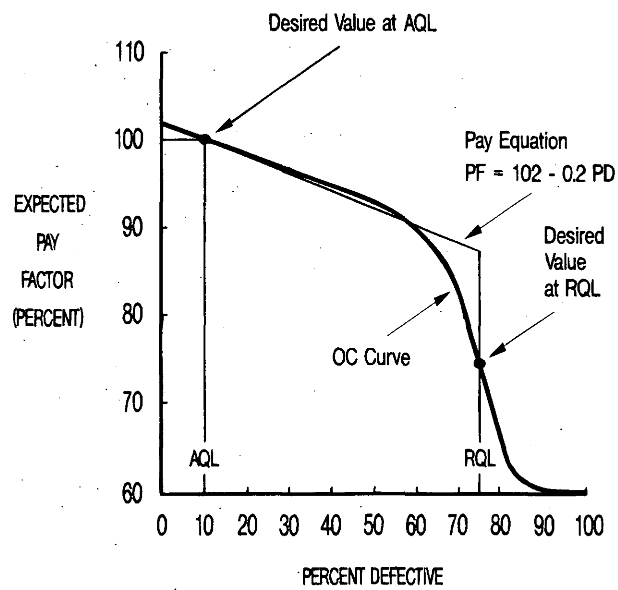


FIGURE 3 OC curve for new air voids acceptance procedure.

The new acceptance procedure is worded as follows in the NJDOT supplemental specifications:

Each mixture in a completed lot shall be compacted so that the combined percentage of material below 2.0 percent voids and above 8.0 percent voids shall be no more than 10 percent. Air voids will be determined from drilled cores taken by the Engineer and tested in accordance with Section 903, Table 903-5 (combined AASHTO/NJDOT procedure). Five cores will be taken at random locations from each lot of approximately 4180 square meters (5,000 square yards) of bituminous concrete of uniform thickness and of approximately 8360 square meters (10,000 square yards) of variable thickness material. Conformance will be judged on the basis of the amount of material estimated to fall outside specification limits as follows:

1. Compute the sample mean ($\bar{X}$) and the standard deviation (S) of the $N = 5$ test results (X_i):

$$\bar{X} = \Sigma X_i / N$$

$$S = [\Sigma (X_i - \bar{X})^2/(N - 1)]^{1/2}$$

2. Compute $Q_L = (\bar{X} - 2.0)/S$ and $Q_u = (8.0 - \bar{X})/S$.
3. Using Table 3, determine the percentage of material falling outside specification limits associated with Q_L and Q_u. Add these two values to obtain the total percent defective, PD.
   a. If PD is less than 50, proceed to Step 5.
   b. If PD is greater than or equal to 50, the NJDOT may elect to reevaluate the lot with additional cores as described in Step 4. If no additional cores are taken, proceed to Step 5.
   c. If PD is greater than or equal to 75, the NJDOT may require the removal and replacement of the defective lot (including any overlying layers) at the Contractor's expense. If this option is not exercised, the Contractor may elect to replace the lot or leave it in place subject to a pay factor of PF = 80 percent.
4. If the NJDOT elects to reevaluate the lot, five additional cores are to be taken at new random locations. Using the five new test results, repeat Steps 1 and 2. Using Table 3 and the computed Q_L and Q_U values, determine the total PD based on the second set of tests. The final PD value is the average of the values obtained from the two sets of tests and is subject to the requirements of Paragraph 3c.
5. Compute the percent payment for the lot as follows:

$$PF = 101 - 0.1\ PD$$

Note that for PD values less than 10, the percent payment exceeds 100, representing a bonus payment.

FIELD TRIALS

The first test of the new specification was a paper exercise applied to several recent jobs that, in the judgment of NJDOT engineers, spanned the range of quality from good to very poor. Because it was the custom with the older specification to assess reductions in terms of material tonnage rather than in terms of a percentage pay factor, the two specifications initially were compared in this manner. The results of this comparison are presented in Table 4.

TABLE 3 Estimation of Percent Defective (PD)

VARIABILITY-UNKNOWN PROCEDURE — SAMPLE SIZE 5 — STANDARD DEVIATION METHOD

Q	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	50.00	49.64	49.29	48.93	48.58	48.22	47.87	47.51	47.15	46.80
0.1	46.44	46.09	45.73	45.38	45.02	44.67	44.31	43.96	43.61	43.25
0.2	42.90	42.54	42.19	41.84	41.48	41.13	40.78	40.43	40.08	39.72
0.3	39.37	39.02	38.67	38.32	37.97	37.62	37.28	36.93	36.58	36.23
0.4	35.88	35.54	35.19	34.85	34.50	34.16	33.81	33.47	33.13	32.78
0.5	32.44	32.10	31.76	31.42	31.08	30.74	30.40	30.06	29.73	29.39
0.6	29.05	28.72	28.39	28.05	27.72	27.39	27.06	26.73	26.40	26.07
0.7	25.74	25.41	25.09	24.76	24.44	24.11	23.79	23.47	23.15	22.83
0.8	22.51	22.19	21.87	21.56	21.24	20.93	20.62	20.31	20.00	19.69
0.9	19.38	19.07	18.77	18.46	18.16	17.86	17.55	17.26	16.96	16.66
1.0	16.36	16.07	15.78	15.48	15.19	14.91	14.62	14.33	14.05	13.76
1.1	13.48	13.20	12.93	12.65	12.37	12.10	11.83	11.56	11.29	11.02
1.2	10.76	10.50	10.23	9.97	9.72	9.46	9.21	8.96	8.71	8.46
1.3	8.21	7.97	7.73	7.49	7.25	7.02	6.79	6.56	6.33	6.10
1.4	5.88	5.66	5.44	5.23	5.02	4.81	4.60	4.39	4.19	3.99
1.5	3.80	3.61	3.42	3.23	3.05	2.87	2.69	2.52	2.35	2.19
1.6	2.03	1.87	1.72	1.57	1.42	1.28	1.15	1.02	0.89	0.77
1.7	0.66	0.55	0.45	0.36	0.27	0.19	0.12	0.06	0.02	0.00

VALUES IN BODY OF TABLE ARE ESTIMATES OF PERCENT DEFECTIVE CORRESPONDING TO SPECIFIC VALUES OF Q = (AVERAGE - LOWER LIMIT) / (STANDARD DEVIATION) OR Q = (UPPER LIMIT - AVERAGE) / (STANDARD DEVIATION). FOR NEGATIVE Q VALUES, THE TABLE VALUES MUST BE SUBTRACTED FROM 100.

TABLE 4 Comparative Performance of New Air Voids Specification

PROJECT	RATING	TONNAGE ADJUSTMENTS OLD SPECIFICATION	NEW SPECIFICATION
1	Good	0	+44
2	Good	0	+27
3	Average	0	-14
4	Average	0	-197
5	Average	-64	-211
6	Poor	-57	-331
7	Poor	-344	-708
8	Poor	-317	-945
9	Very Poor	-1324	-1613

TABLE 5 Data from Pilot Projects with New Air Voids Specification

LOT	TYPE	VOIDS CONTENT (PERCENT) AVERAGE	STANDARD DEVIATION	PAY FACTOR (PERCENT)
Project #1:				
1	Surface	8.06	1.61	95.9
	Base	5.66	0.75	101.0
2	Surface	6.88	1.43	98.7
	Base	5.34	0.61	101.0
3	Surface	7.56	1.68	96.9
	Base	5.26	0.98	101.0
4	Surface	7.22	1.54	97.8
	Base	5.74	1.36	100.9
5	Surface	7.04	1.79	97.9
	Base	4.38	0.69	101.0
6	Surface	7.12	2.52	97.2
	Base	5.72	1.17	101.0
7	Surface	8.06	2.09	95.1
	Base	5.72	1.18	101.0
8	Surface	7.30	2.13	97.2
	Base	4.92	0.89	101.0
9	Surface	6.26	0.82	101.0
	Base	5.06	1.02	101.0
10	Surface	9.66	2.44	93.6
	Base	5.48	0.79	101.0
11	Surface	5.90	1.14	101.0
	Base	5.32	0.73	101.0
12	Surface	8.10	1.62	95.8
	Base	6.74	1.59	98.7
Project #2:				
1	Surface	6.12	1.38	100.4
	Base	5.16	0.62	101.0
2	Surface	8.00	4.12	95.5
	Base	5.48	0.87	101.0
3	Surface	9.38	3.01	94.4

The following observations can be made from the data in Table 4. The two jobs judged to be of good quality would have received small bonuses under the new specification; the three regarded as average would have received some amount of pay reduction; the four considered to be either poor or very poor would have received substantially greater pay reductions than under the older specification. This result was both expected and desired and, on the basis of these results, it was decided to proceed with actual field trials.

Two jobs were selected to serve as pilot projects. They were chosen to be as nearly representative as possible of typical construction, they were constructed by different contractors in different geographic areas of the state, and they were completed during the regular construction season. The data, representing a total of 29 lots of both base and surface courses, are presented in Table 5.

It can be seen from the data in Table 5 that the air voids levels in the base layers were well controlled but that some difficulty was encountered with the surface layers. This situation has often been the case and was part of the motivation for revising the specification.

There are several possible reasons for this difference in level of compaction. One is that the base layers generally benefit from the additional compactive effort applied to the surface layers. Another possible reason is that because of the normal sequence of events as the job is constructed, the surface layers are sometimes placed later in the year when cooler weather makes compaction more difficult. It has also been suggested that, because the base layers are often placed on underlying layers with rougher surface texture, there may be less slippage, which results in more efficient roller action. Finally, roller patterns that have been established to be effective when the base layers are placed may not necessarily be optimal when the surface layers are placed.

IMPLEMENTATION

It was concluded from the experience on the two pilot projects that the requirements of the specification could be achieved but that an additional trial period would be beneficial. Consequently, it was decided to implement the new specification on all future projects but to retain the relaxed pay equation and RQL requirements for at least another construction season.

Several additional projects have been completed and the overall performance continues to improve. In the few cases in which problems have been encountered, it has usually been possible to identify a cause. Clearly, the specification makes it incumbent on the contractor to pay close attention to quality control, good construction practices, and the appropriate design level for air voids. Another evaluation will be made after more experience is gained with this specification, and a decision will then be made regarding future refinements.

SUMMARY

The NJDOT specification for air voids in bituminous concrete has been upgraded in accordance with the following objectives:

- The specification is to be performed-based.
- It must provide sufficient incentive to produce high-quality work.
- It must be technically sound and fair to all parties.
- It must be practical and administratively efficient.

The new specification is performance-related in that it controls air voids, a surrogate measure of compaction, which is highly correlated with service life. The pay schedule, which awards bonus payments for superior quality and assesses pay reductions for deficient quality, is believed to provide ample incentive to produce good work. The procedure is technically sound in that it uses efficient statistical measures in a correct way. OC curves were constructed to verify that the procedure fairly provides 100 percent payment at the AQL and withholds sufficient payment at lower quality levels to cover the anticipated costs of future repairs. It is practical in that the acceptance procedure is easy to understand and administer.

Two pilot projects were completed successfully, and the new specification was adopted for all future projects using an interim form of the pay schedule. A continuing evaluation will determine when and if additional modifications should be made.

REFERENCES

1. *Federal Register,* Vol. 49, No. 115, June 13, 1984, p. 24375.
2. 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.
3. Linden, R. N., J. P. Mahoney, and N. C. Jackson. Effect of Compaction on Asphalt Concrete Performance. In *Transportation Research Record 1217,* TRB, National Research Council, Washington, D.C., 1989.
4. *Quality Management.* Demonstration Project 89. Demonstration Projects Division, FHWA, U.S. Department of Transportation, 1993.

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

Smoothness Control in Asphalt Pavement Construction: Development of Specifications, Implementation, and Results

MUSTAQUE HOSSAIN AND WILLIAM H. PARCELLS, JR.

Surface smoothness on newly constructed pavement is a major concern of the highway industry. This smoothness, or riding comfort, is an indication of the quality of the newly constructed pavements since it affects road users directly. Smoothness specifications for asphalt concrete (AC) pavements now in effect in Kansas have evolved over the past few years through a number of revisions. Pavement profiles with short wavelengths and smaller amplitudes than the industry-accepted 5.1 mm (0.2 in.) can harm the ride quality of pavements. This experience has led the Kansas Department of Transportation to eliminate the blanking band width in the profilograph trace reduction process. The implementation of this zero, or null, blanking band was successful and has resulted in smoother asphalt pavements in Kansas. The currently used specifications for AC pavements have been based on the consideration of a number of factors related to the construction of and the measurement of smoothness on AC pavement. The incentive payment amounts have been calculated to make these compatible with the incentive payments for concrete pavement. The results show that these smoothness specifications can be achieved by contractors, and the number of sections in the bonus range indicates that the incentive payments encouraged better-quality paving. These results should have a positive impact on AC paving in Kansas. Overall, an increasing number of miles of pavements with lower profile index values are being constructed since the implementation of smoothness specifications for AC pavements.

Pavement smoothness and roughness can be described by the magnitude of profile irregularities and their distribution over the measurement interval. The road surface smoothness on newly constructed pavement is a major concern of the highway industry. This smoothness, or riding comfort, is a measure of the quality of the newly constructed pavements since it affects the road users directly. According to Hudson (*1*), the primary purpose for smoothness measurement is to maintain construction quality control.

It is accepted that there is a growing interest in the highway industry for attaining smoother and smoother pavement surfaces. Results from a 1992 NCHRP study show that of the 22 states reporting, 91 percent used smoothness criteria on new pavement construction (*2*). In 1990 NCHRP reported that of 36 states reviewed, 80 percent utilized smoothness criteria on new pavement construction (*3*). The increasing trend in the use of ride quality specifications is also evidenced by the 1992 study, in which 21 states out of 25 queried believe that there will also be a future increase in ride quality requirements. A 1987 AASHTO survey showed that 53 percent of the states using profilographs for acceptance of concrete pavements used incentive and disincentive specifications (*4*). The incentive and disincentive values in smoothness specifications typically ranged from 1 to 5 percent of the bid item price, with 31 percent of the states reporting allowable incentives up to 5 percent. The relatively high incentives now possible with many of the profilograph specifications place an ever-increasing burden on the measurement process and data reduction process. Variability in test results can substantially affect contractor payments (*2*).

Mustaque Hossain, Department of Civil Engineering, Kansas State University, Manhattan, Kan. 66506. W. H. Parcells, Kansas Department of Transportation, Bureau of Materials and Research, Materials and Research Center, 2300 Van Buren, Topeka, Kan. 66611.

DEVELOPMENT OF AC PAVEMENT SMOOTHNESS SPECIFICATIONS

Factors Considered

In 1985 the Kansas Department of Transportation (KDOT) selected a 7.63-m (25-ft) California-type profilograph and a 5.1-mm (0.2-in.) blanking band for evaluation of the profilogram for determining the smoothness of portland cement concrete (PCC) pavement construction (*5*). In 1985 the first three PCC pavement projects with smoothness requirements were constructed. However, the incentive clauses were not exercised. Profilograph measurements were taken on each wheel path. The profilograph results in terms of profile roughness index (PRI) on 0.16-km (0.1-mi) intervals on these projects were analyzed. The first two projects had a high percentage of sections in the bonus range indicating that smoothness of 0 to 63 mm/km (0 to 4 in./mi) was practical and achievable. In 1990 the specifications given in Table 1 were adopted as standards for controlling concrete pavement smoothness in Kansas.

Although smoothness specifications with profilograph measurements were implemented on PCC pavements in 1985, new bituminous pavements had surface tolerance requirements as measured by a 3.05-m (10-ft) straight edge or a 7.62-m (25-ft) stringline at selected locations. The maximum variation of the surface in 3.05 m (10 ft) was not allowed to exceed 4.76 mm (3/16 in.); the maximum for 7.62 m (25 ft) was 7.94 mm (5/16 in.) (*6*). Evidently these requirements were not sufficient for constructing smooth-riding bituminous pavements, and public complaints about the quality of rides on newly paved asphalt concrete (AC) pavements were rampant. By 1990 KDOT was very successful in controlling concrete pavement smoothness. This success led to the development of profilograph-based specifications for AC pavements in 1990. The major elements of the smoothness specifications for asphalt pavements evolved through consideration of the following:

• The roadway elements that normally would be included in the smoothness specifications for bituminous pavements are finished surfaces of the mainline pavement, side roads, auxiliary lanes, and ramps. Each of these elements should have a minimum paving depth of 102 mm (4 in.). This minimum thickness was selected because of economics. All of KDOT's substantial maintenance projects have actions that are less than 102 mm (4 in.) thick. Because of budget restraints on the substantial maintenance program money, it was believed that this money should not be spent on incentives.

• Unlike concrete pavements, there are no hand-poured sections on any of the elements described previously. Thus, a single set of specifications would be developed irrespective of the posted speed limit on the roadway.

• Specifications would be developed for statewide application regardless of route type or functional classification. This should encourage the contractors to pave uniformly throughout the state.

• The following would be excluded from pay adjustments under the terms of the smoothness specifications:

–Bridge decks unless to be overlaid,

–Acceleration and deceleration lanes for at-grade intersections,

–Shoulders,

–Pavement on horizontal curves that have a 304.8-m (1,000-ft) or less centerline radius of curvature and pavement within the superelevation transition of such curves,

–Pavements consisting of new or recycled bituminous concrete surfacing 102 mm (4 in.) or less in plan thickness,

–County secondary and federal aid urban projects unless specified otherwise on the plans, and

–Projects less than 0.5 mi in length (excluding bridge lengths).

• The California-type profilograph would be used for as-constructed smoothness measurements, and the schedule for adjusted payment would be fashioned after that for concrete pavements. Doing this will bring some kind of parity between the specifications for these competing types of pavements. It was accepted that during paving of bituminous pavements, contractors had a better opportunity to meet smoothness requirements than during paving of concrete pavements, so the disincentive payments would be much higher for bituminous pavements. The specification for the "bumps" would be similar to that for concrete pavements (deviations in excess of 10.2 mm in a length of no more than 7.6 m, or 0.4 in. in 25 ft).

• The pavement smoothness would be established as a separate pay item with a zero-bid item amount. The pay schedule would then include incentives or disincentives in accordance with the pay schedule that will be added to or subtracted from the total contract amount through this pay item.

TABLE 1 Schedule for Adjusted Payment for PCC Pavements, 1990 Specification 502.06 with 5.1-mm or 0.2-in. Blanking Band

Profile Index millimeter per kilometer per 0.16 kilometer segment	Price Adjustment Percent of Contract unit bid price
48 or less	106
48.1 to 64	103
64.1 to 159	100
159.1 to 191	96
191.1 to 222	92
222.1 to 238	90
238.1 or more	88 (Corrective Work required or replace)

TABLE 2 Schedule for Adjusted Payment for AC Pavements, Special Provision 90P-39 with 5.1-mm or 0.2-in. Blanking Band

Profile Index (mm per km per 0.16 lane-km)	Contract Price Adjustment per 0.16 lane-km (Dollars)
32 or less	+ 152.00
32.1 to 47	+ 76.00
47.1 to 142	0.00
142.1 to 174	-102.00
174.1 to 205	-203.00
205.1 to 237	-254.00
237.1 or greater	-305.00

• The incentive or disincentive amounts would be determined on a 0.16-km (0.1-mi) basis, which would be summed to an aggregate amount on a lane-mile basis. There should be a threshold target that when exceeded would require specific remedial action by the contractor such as that in the concrete pavement specification (i.e., grinding).

• The test method and trace reduction procedures would be similar to those used for concrete pavements.

• The contractor would be responsible for determining the smoothness of pavement by operating a profilograph. KDOT may perform profilograph testing on the surface for monitoring and comparison purposes and during disputes over test results.

On the basis of these considerations, profilograph results for ensuring smoothness on bituminous pavements with greater than 102 mm (4 in.) paving depth were implemented through Special Provision 90P-39 as given in Table 2, where the roughness limits established were somewhat similar to the PCC pavement schedule.

Calculation of Incentive Payments

The highest incentive payment of $152/0.16-km (0.1-mi) section as presented in Table 2 for the profile index of 32 mm/km (2.0 in./mi) or less was based on the average cost of an AC overlay 89 mm (3.5 in.) thick, 161 m (528 ft) long, and 3.66 m (12 ft) wide. Many states pay for AC paving by the square yard paved; KDOT pays for AC paving by the tons of mix and again for the asphalt binder. For the Kansas condition, there was no direct correlation between pay items for AC and PCC pavements. Therefore, a direct conversion of PCC incentive payments for smoothness to AC condition was not possible. The 89-mm (3.5-in.) AC pavement thickness was arbitrarily selected because it was thought that the incentive payment should be compared with that for PCC pavement at this thickness level. Details of this incentive payment calculation are as follows:

• The amount of AC in an overlay section 89 mm (3.5 in.) thick, 161 m (528 ft) long, and 3.66 m (12 ft) wide (unit weight = 145 pcf):

$0.2917 \times 145 \times 12 \times 528/2{,}000 = 134$ tons

• The cost of this section of overlay (based on the price of AC, BM-2 for KDOT):

134 tons $\times$ \$18.90/ton = \$2,532

- The price of this section including incentive payment (maximum 106 percent, based on then-current PCC pavement payment adjustment schedule):

$2,532 × 1.06 = $2,684

- The maximum amount of incentive for a 0.16-km (0.1-mil) section =

$2,684 − $2,532 = $152.00

The payment schedule for the profile/index 33.1 to 47 mm/km (2.1 to 3.0 in.) was established to be half the amount for 0 to 32 mm/km (0 to 2.0 in./mi) (i.e., $76/0.16-km or 0.1-mi section). The disincentive amounts were made progressively higher (up to $305/0.16-km or 0.1-mi section) to discourage contractor negligence.

IMPLEMENTATION OF AC PAVEMENT SMOOTHNESS SPECIFICATIONS

As mentioned earlier, the profilograph results for ensuring smoothness on AC pavements with greater than 102-mm (4-in.) paving depth were implemented through Special Provisions 90P-39 (as given in Table 2) during 1990. During this year, the incentive and disincentive clauses were not enforced. The profilograph results were collected and analyzed using the 5.1-mm (0.2-in.) blanking band. Table 3 gives the specification compliance for the 5.1-mm (0.2-in.) blanking band. Out of 851 sections (0.16-km or 0.1-mi) constructed in 1990, there were 547 sections (64 percent) in the bonus range, 226 sections (27 percent) in the full-pay range, and 78 sections (9 percent) in the penalty zone. Figure 1 illustrates the results; no specific statistical distribution is obvious. Most of the sections were lumped in the bonus range. However, the data contained some sections on which profilograph specifications were not required but were considered rough and measurements were made.

TABLE 3 Profilograph Results on AC Pavements Using 5.1-mm Blanking Band for 1990 Special Provision 90P-39

Roadway	No. of 0.16 kilometer sections	Compliance with specified PRI (mm/km)					
		PRI (0 - 47) Bonus	(%)	PRI (47.1 - 142) Full-pay	(%)	PRI (>142) Penalty	(%)
All 1990	851	547	64	226	27	78	9

REVISED PROFILOGRAPH TRACE REDUCTION PROCEDURE

In 1990 there was a noticeable high-frequency vibration on a concrete pavement reconstruction project on I-70. This vibration was not noticed for another concurrent new PCC pavement project on I-470, however. A closer review of the profilograph traces on these projects showed that on the I-70 project, there was a consistent sine-wave cyclic oscillation of about 2.44-m (8-ft) spacing and with 5.1-mm (0.2-in.) amplitudes. Most of these surface deviations were covered up by the 5.1-mm (0.2-in.) blanking band during trace reduction. On the I-470 project, the oscillation waves were of about 9.14-m (30-ft) spacing and about 5.1-mm (0.2-in.) amplitude, which were, again, covered up by the 5.1-mm (0.2-in.) blanking band during trace reduction (*7*). This issue of the effects of short wavelengths on PRI was tied to the question about the proper blanking band width.

The I-70 and I-470 projects of 1990 prompted KDOT to experiment with the blanking band width in order to quantify the apparent visual difference of profilograph traces on these projects. It was decided to use a zero blanking band width, or null blanking band. The null blanking band is nothing but a reference line usually placed approximately at the center of the trace having the line equally dividing the scallops above or below the centerline. The null blanking band was also extended to cover profilograms from bituminous pavements.

Reanalysis of the profilograms from the AC pavement projects of 1990 was done using the null blanking band. Table 4 presents the

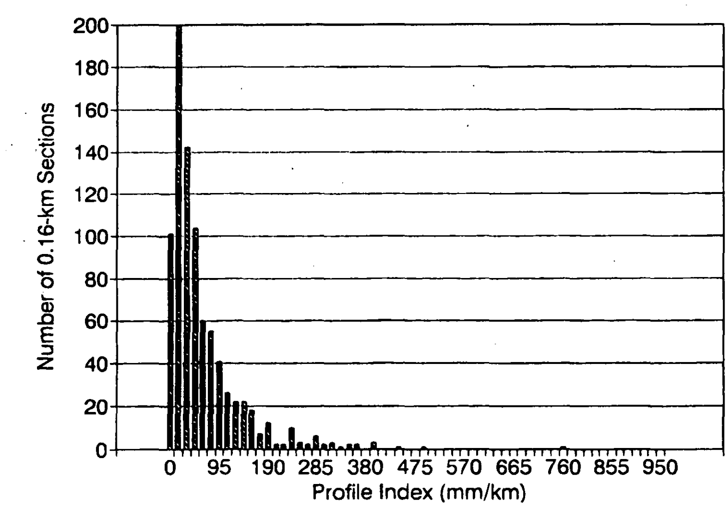


FIGURE 1 Specification compliance of AC pavement sections with Special Provision 90P-39 (0.2-in. or 5.1-mm blanking band).

TABLE 4 Profilograph Results on AC Pavements Using Null (0.254-mm) Blanking Band for 1990 Special Provision 90P-39

Roadway	No. of 0.16 kilometer sections	Compliance with specified PRI (mm/km)					
		PRI (0 - 158) Bonus	(%)	PRI (158.1 - 631) Full-pay	(%)	PRI (>631) Penalty	(%)
All 1990	842	71	8	753	90	18	2

specification compliance for the null blanking band. Out of 842 sections analyzed, 71 sections (8 percent) were in the bonus range, 753 sections (90 percent) in the full-pay range, and 18 sections (2 percent) in the penalty range; Figure 2 illustrates the results. The distribution of the measurements is somewhat normal, which should be expected for a set of engineering measurements. It appears that the null blanking band has enhanced the ability of the profilograph to measure the smoothness of newly constructed AC pavements. However, these results made it obvious that the specifications of Special Provision 90P-39 needed to be changed in order to interpret null blanking band results (*5*).

REVISION OF AC PAVEMENT SMOOTHNESS SPECIFICATIONS

In 1991 Special Provision 90P-39-R1 was incorporated for AC pavement projects that also required the use of the null blanking band for mechanical profilographs or 0.25-mm (0.01-in.) blanking band for computerized profilographs. The corrective action for a rough section was modified slightly, eliminating the requirement to reseal the diamond-ground pavement, and incorporated in Special Provision 90P-39-R2 in 1992. The schedule for adjusted payments in this special provision at various levels of smoothness achieved in construction is given in Table 5. This requirement was applicable to all projects with multiple paver passes including cold milling with overlay or cold recycle with an overlay. The working depth in those cases might be less than 102 mm (4 in.). However, pay adjustment did not apply if the plan thickness is less than 102 mm (4 in.) on the existing surfaces (*6*).

In 1993 the results of profilograph testing on 5,866 0.16-km (0.1-mi) sections from 30 paving projects in 1992 were analyzed using the null blanking band and compared with the results from the sections of 1990 and 1991. Table 6 gives the results. The 1992 results showed an increased percentage of sections in the bonus range with a similar reduction in the full-pay group. It is apparent that the smoothness limits in Special Provision 90P-39-R2 were achievable (*8*).

During the implementation of Special Provision 90P-39-R2, some contractors complained that requiring all pavement sections to be profiled on the same day they were placed was causing the contractors to stop paving earlier during afternoon hours in order to have time to finish rolling and profiling before reopening the highway to traffic.

Special Provision 90P-39-R3 contains an option allowing the contractor to delay profiling the final portion of a day's paving (not to exceed five 0.16-km or 0.1-mi sections) until the first working day that production is continued on the same lane. When deciding whether to exercise this option, the contractor should be aware that the profile index of the pavement will probably be higher after it has been opened to traffic than it would have been if profiled as soon as rolling was completed.

As more and more AC pavement projects were being built with these smoothness specifications, the clauses of grind-back provisions to profile index of 394 mm/km (25 in./mi) or less in Special Provision 90P-39-R3 were disputed by the contractors. They argued that if they had achieved a profile index of 473 mm/km (30.0 in./mi) then no grinding would have been necessary but that a profile index of 473.1 mm/km (30.1 in./mi) would require grinding back to 394 mm/km (25 in./mi). Special Provision 90P-39-R4 and subsequent revision 90P-39-R5 now require grind-back to 473 mm/km (30.0 in./mi) or less in case of a measured profile index greater than 473 mm/km (30.0 in./mi) along with the penalty payment, if any.

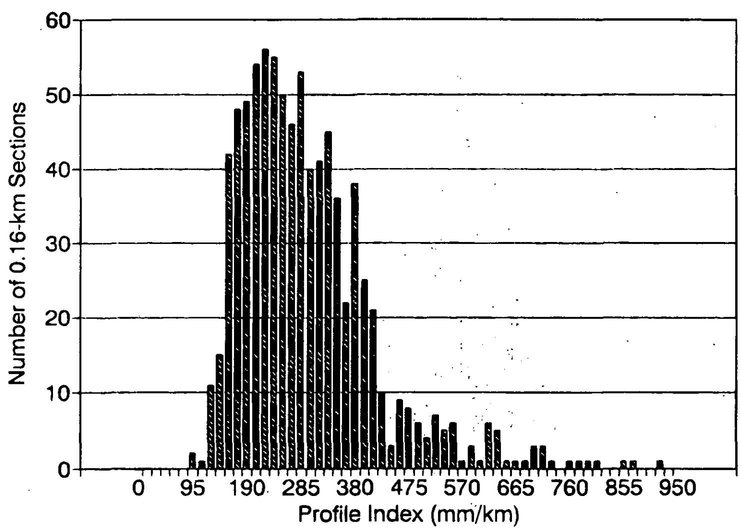


FIGURE 2 Specification compliance of AC pavement sections with Special Provision 90P-39 (null blanking band).

TABLE 5 Schedule for Adjusted Payment for AC Pavements, Special Provision 90P-39-R2

Profile Index (mm per km per 0.16 lane-km)	Contract Price Adjustment Per 0.16 Lane-km (Dollars)
110 or less	+ 152.00
110.1 to 158	+ 76.00
158.1 to 473	0.00
473.1 to 631	0.00 (correct back to 394 mm/km or less)
631.1 to more	-203.00 (correct back to 394 mm/km or less)

TABLE 6 Profilograph Results on AC Pavements Using Zero (0.254-mm) Blanking Band for 1993 Special Provision 90P-39-R2

Roadway	No. of 0.16 kilometer sections	Compliance with specified PRI (mm/km)					
		PRI (0 - 158) Bonus	(%)	PRI (158.1 - 631) Full-pay	(%)	PRI (>631) Penalty	(%)
1990 (reanalysis)	842	71	8	753	90	18	2
1991 (reanalysis)	1890	57	3	1796	95	37	2
1992	5866	1467	25	4341	74	58	1
1993	4166	625	15	3499	84	42	1

CURRENT SITUATION

In 1994 the results of profilograph measurements on 4,166 sections of AC pavement were collected from 24 paving projects completed in 1993. Table 6 presents the trace reduction results. Figure 3 illustrates the results graphically, and a normal distribution of the results is apparent. The traces were reduced using a null blanking band and the results were compared with those of 1990, 1991, and 1992. There is a decreased percentage of sections in the bonus range with a similar increase in the full-pay group (*9*). The results show that although the incentive payments have decreased, the currently used specifications for AC pavements are achievable by the contractors. The results should establish that under current specifications, bonus can be achieved through better-quality paving rather than by chance.

COST ANALYSIS OF AC PAVEMENT SMOOTHNESS SPECIFICATIONS

The incentive and disincentive payments made to the contractors in 1991, 1992, and 1993 were analyzed to determine a trend in such payments. Table 7 provides the results of this analysis, and Figure 4 illustrates the results graphically. The incentive payments were much higher during the second year of the implementation of smoothness specification. The incentive to lane-kilometer-paved ratios were 14.07, 203.59, and 129.25 (22.5, 325.7, and 206.8 for lane miles) for 1991, 1992, and 1993, respectively. As illustrated in Figure 4, the incentive payments were showing trends of stabilization after a sharp increase in 1992. The disincentive payments were very minimal compared with the incentive payments in 1992 and 1993; they were somewhat stable then, also. However, the opposite was true in 1991. This indicates that the new specifications have made a positive impact on the overall quality of AC paving over the past 3 years.

TABLE 7 Results of Cost Analysis of AC Pavement Smoothness Specifications

Year	No. of 0.16-km Sections	Bonus ($)	Bonus/ Lane-km Paved ($/km)	Penalty ($)	Penalty/ Lane-km Paved ($/km)
1991	1890	4256	14.07	7919	26.19
1992	5866	191084	203.59	4060	13.43
1993	4568	94468	129.25	3857	12.75

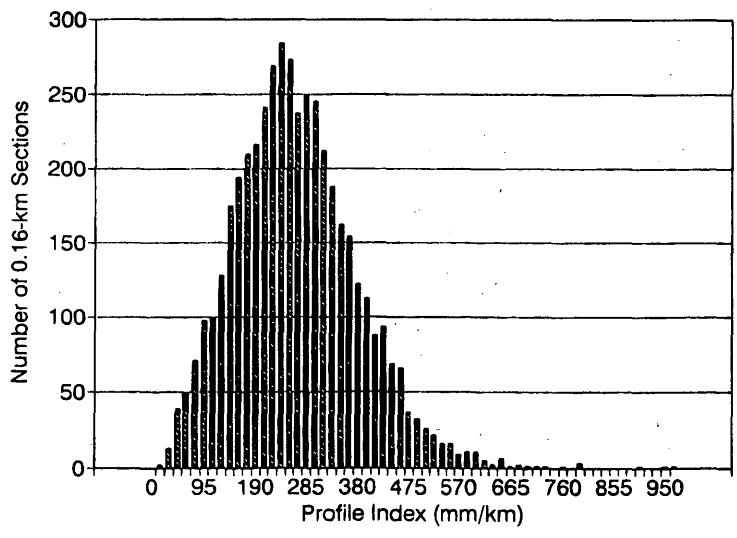


FIGURE 3 Specification compliance of AC pavement sections with Special Provision 90P-39-R2 (null blanking band).

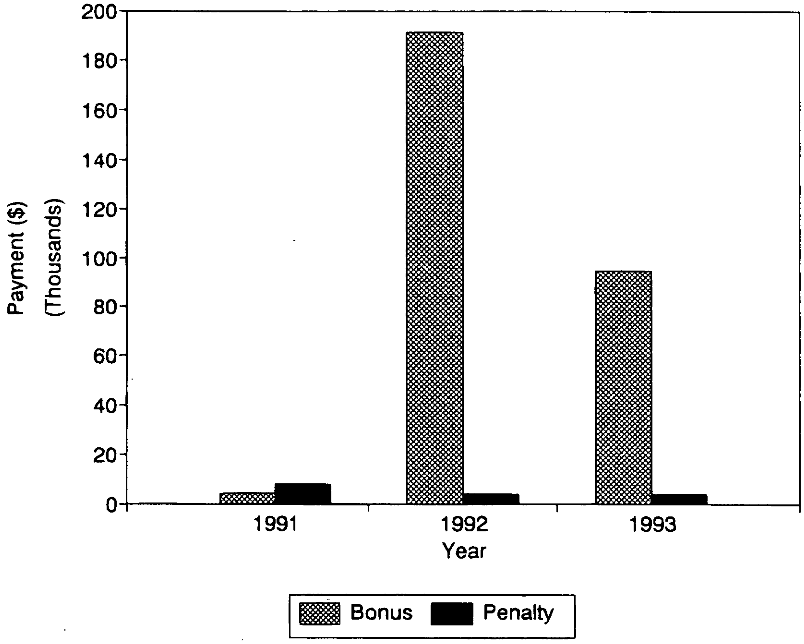


FIGURE 4 Incentive and Disincentive Payments for AC Paving in 1991, 1992, and 1993.

CONCLUSIONS

Smoothness specifications for AC pavements now in effect in Kansas have evolved over the past few years through a number of revisions. Pavement profiles with short wavelengths and smaller amplitudes than the industry-accepted 5.1 mm (0.2 in.) can harm the ride quality of pavements. This experience has led KDOT to eliminate the blanking band width in the profilograph trace reduction process first for concrete pavements, then for bituminous pavements. The implementation of this zero, or null, blanking band was successful and has resulted in better-quality pavements in Kansas. The currently used specifications for AC pavements can be achieved by contractors, and the number of sections in the bonus range indicates that incentive payments have encouraged better paving than in the past. This should have a positive impact on asphalt pavement paving in Kansas. In general, an increasing number of miles of pavement with low profile index are now being constructed since smoothness specifications for bituminous pavements were implemented.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support provided by KDOT for this study. Thanks are due to Ron Harvey, Construction Engineer of Bureau of Construction and Maintenance; Lon Ingram, Chief of Bureau of Materials and Research; and Dick McReynolds, Engineer of Research, for their continued interest in and support of smoothness research.

REFERENCES

1. Hudson, W. R. Road Roughness: Its Elements and Measurement. In *Transportation Research Record 836,* TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
2. Scofield, L. A. *Profilograph Limitations, Correlation, and Calibration Criteria for Effective Performance-Based Specifications.* Final Report, NCHRP Project 20-7, Task 53. TRB, National Research Council, Washington, D.C., Dec. 1992.
3. Woodstrom, J. H. *NCHRP Report 167: Measurement, Specifications, and Achievement of Smoothness for Pavement Construction.* TRB, National Research Council, Washington, D.C., Nov. 1990.
4. *Summary of Results of the 1987 AASHTO Rideability Survey.* Highway Subcommittee on Construction, AASHTO, Washington, D.C., 1988.
5. Parcells, W. H., Jr. *Control of Pavement Trueness in Kansas.* Interim Report. Kansas Department of Transportation, Topeka, Jan. 1992.
6. *Standard Specifications for State Road and Bridge Construction.* Kansas Department of Transportation, Topeka, 1990.
7. Parcells, W. H., Jr., and M. Hossain. Kansas Experience with Smoothness Specifications for Concrete Pavements. In *Transportation Research Record 1435,* TRB, National Research Council, Washington, D.C., 1994, pp. 115–123.
8. Parcells, W. H., Jr. *Control of Pavement Trueness in Kansas.* Interim Report. Kansas Department of Transportation, Topeka, May 1992.
9. Parcells, W. H., Jr. *Control of Pavement Trueness in Kansas.* Interim Report. Kansas Department of Transportation, Topeka, May 1994.

Publication of this paper sponsored by Committee on Flexible Pavement Construction and Rehabilitation.

PART 2

Construction Management

Decision Support System Framework for Construction Technology Transfer and Diffusion

MOHAMED-ASEM U. ABDUL-MALAK, TOUFIC MEZHER, AND E. LILE MURPHREE, JR.

Advanced construction technologies have emerged over the past decade that cover a wide range of new applications, particularly for the equipment-intensive industry of highway construction. However, the reluctance of highway contractors to implement these technologies has caused their slow diffusion in global construction markets. The thrusts for construction technology transfer and diffusion are identified as are the factors that may impede the transfer and diffusion process. A framework for decision making that incorporates the identified factors is then proposed; contractors can use the framework to evaluate the feasibility of adopting advanced construction technologies. The proposed framework uses the knowledge available on emerging technologies and guides the decision maker into either a rule-based analysis of potential barriers to the technology transfer and diffusion process or an analytic hierarchy process evaluation of factors that promote and impede technology, depending on the perceived level of risk exposure.

The highway construction industry is characterized by its dependency on heavy equipment as the applied resource most vital to construction work. In addition, progress at highway construction sites generally is paced by the output of the equipment and the interdependencies among the construction operations. Emerging technologies for the highway construction industry have been aimed at increasing the productivity rates and efficiency of the equipment used. This increase usually is due to automated processes that rely on sensing and microprocessing technologies. However, the decision to acquire such technologies may be accompanied by a high initial investment requirement and may necessitate a certain level of work force skills needed to safely operate and maintain the acquired systems. Other technical constraints such as governmental regulations, outdated project specifications, and project site conditions may render a technology transfer decision impractical.

In this research, factors that may urge the transfer and diffusion of emerging technologies, as well as those that may act to slow, delay, or impede the process, have been identified. The paper presents a decision support system framework that incorporates the factors in two alternative analysis approaches: (*a*) a rule-based screening of potential barriers and (*b*) an analytic hierarchy process (AHP) evaluation. The proposed structure of the decision process is intended to help highway contractors visualize how an individual decision determinant or group of determinants may cause a variation in the final decision.

M.-A. U. Abdul-Malak and T. Mezher, Faculty of Engineering and Architecture, American University of Beirut, Beirut, Lebanon. E. L. Murphree, Jr., Department of Engineering Management, School of Engineering and Applied Science, George Washington University, Melvin Gelman Library, Room 632, Washington, D.C. 20052.

DECISION ENVIRONMENT

Highway contractors operating in highly competitive construction markets must decide the feasibility and timing of investing in emerging construction technologies. The decision environment requires that the drives favoring technological change be weighed against the possible impediments. Although the drives are generally technology-dependent, the impediments are most likely to be related to (*a*) the type and characteristics of construction projects; (*b*) the size, strength, and resources of the construction firm; (*c*) the practices and policies of highway authorities and agencies; and (*d*) the government regulations imposed in the area of work. The final decision, however, may vary depending on how different contractors perceive the risk exposure caused by impediments and on the prevailing condition of the overall construction market. Figure 1 is a schematic representation of the environment within which technology transfer and diffusion decisions are made.

Thrusts for Technology Transfer and Diffusion

Highway contractors favor advanced technologies mainly for the competitive advantage that such technologies offer, both at the domestic and international levels (*1,2*). This advantage may be gained in one or more of the three possible forms: reduced bid prices (*3,4*), shorter construction schedules (*5*), and higher levels of achieved quality (*6*). Although the hourly ownership and operating cost—based on the required initial investment cost and the estimated operation and maintenance costs prorated over the expected life of a technology—may be higher for the advanced technology under consideration, the cost per unit of work may be less than that offered by the conventional technology. This difference is attributed to the higher production rate of the advanced technology and the resulting shorter duration required to accomplish the work. In addition, if the operation that uses the new technology is one of those most likely to be critical, a reduced overall project duration may be realized. Completing a project sooner could be a major advantage to highway contractors if project schedule is a parameter in the bidding evaluation process. Similarly, as new technologies have been aimed at improving the quality of constructed facilities—particularly when such improvement can be translated into lower facility life-cycle costs—competitive advantage can be gained by contractors with a multiparameter bidding system.

Sometimes, new technologies may solve technical problems that can only partially be overcome with conventional technologies. For example, in texturing an existing concrete pavement surface to develop a good bond with a new asphalt concrete layer, the use of

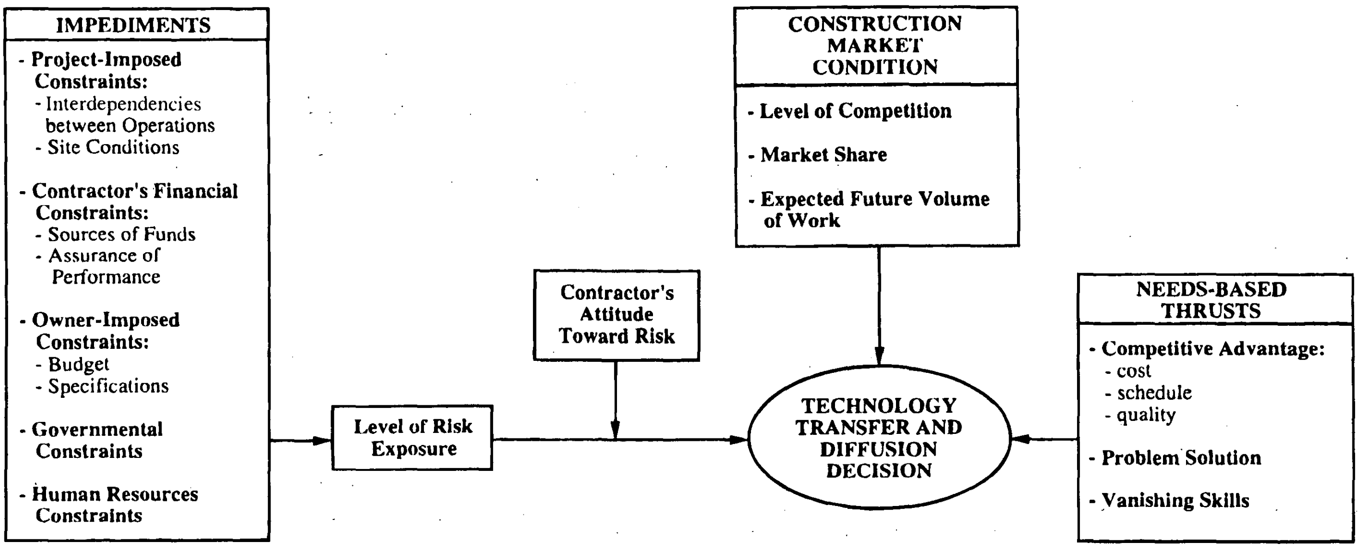


FIGURE 1 Schematic representation of factors affecting technology transfer and diffusion decisions.

a shotblasting robot equipped with automated visual-sensing capabilities could help to ensure a more uniform concrete texture, thus reducing the likelihood of delamination between the old and new surfaces.

As discussed, a new technology may be geared to reduce the level of labor skills needed to achieve the precision requirements specified (*1,6*). Another example involves the use of a laser-based grader whose blade can be operated in either a fully automated or a semi-automated mode. Such a technology requires little or no operator involvement in initiating an adjustment in the height of the grader's blade, but its use would result in a tremendous productivity gain compared with that of conventional grading, particularly in relation to the level-of-skill input required of the operator.

Impediments to Technology Transfer and Diffusion

Many factors may contribute to rendering the use of an advanced technology unfeasible. Constraints imposed by a project's characteristics may be in the form of an interdependency between two or more construction operations or a physical condition at the construction site. A progress-based relationship between any two project activities is an example of the former constraint, whereby the introduction of an advanced technology to one of the operations may not yield the full benefit intended. That is, shortening the duration of the activity in question in a way that uses the maximum production rate of the incorporated technology may not be possible because of the progress-based relationship with the other activity, for which a compatible, more productive technology has not yet emerged. The latter constraint can be exemplified by a permissible limit of longitudinal grade, among other geometric features, beyond which the higher level of performance of the laser-based grader will be jeopardized.

Other types of project-related constraints are those imposed by the owner (the highway agency). If the primary concern of highway agencies is to award projects on the basis of the lowest bid, technologies that offer marginal schedule and quality benefits but not a reduced bid price probably will be deemed unfeasible by highway contractors, because using such technologies may lower their chances of winning project contracts. Higher bid prices may be the result of a low projected volume of work for which the technology could be used or an additional increment in the initial equipment investment possibly due to high taxes imposed by the government on imported technologies. Another project-related, owner-imposed constraint is the method used by highway agencies to specify execution requirements. Although the performance method of specifying is thought to promote the use of advanced technologies, the descriptive method, when specifying outdated requirements, can severely hinder the application of newer techniques.

Constraints that pertain to a contractor's financial strength can be related to a contractor's ability to secure the funds necessary for acquiring a new technology and to bond the contract against potential performance defaults while experimenting with the new technology. With higher financing and bonding premiums, the possibility is greater that the resulting bid prices will be less competitive. On the other hand, the unfamiliarity of contractors' labor resources with the new technology and their inability to efficiently operate and maintain it pose another setback that could lead to performance defaults and financial losses.

Government constraints may be of two main forms: moderate or strict regulations. Examples include high taxes on imported technologies and bans on the import of such technologies, respectively. Such control to protect could be intended to protect the domestic equipment manufacturing industry or to protect the interest of the local labor-intensive economy, particularly in cases where imported technologies are expected to reduce the labor requirements on construction sites.

Construction Market Condition

Final decisions on new technology made by contractors operating in the same construction market may still vary depending on how each contractor perceives the level of risk exposure involved. In addition, the level of competition, a contractor's share of the market, and the projected volume of work to which a technology can be applied all contribute to either promoting or hindering the technology implementation process (*7*). For example, laser-based grading technology would probably be feasible in countries where the volume of new highway construction work is anticipated to be large. On the other hand, it may not be of interest to highway contractors in areas where the highway networks have matured and where highway agencies would emphasize maintaining and rehabilitating the existing networks.

DECISION-MAKING FRAMEWORK

Existing frameworks of technology-transfer decisions are based on the identification of critical factors affecting the decision process. Building codes, conservatism, and organizational barriers are reported to be major determinants in building construction transfer decisions (*8*). Two organizational approaches—top down and bottom up—have been identified to delineate the possible paths for technology transfer (*9*). In these approaches, the transfer process is shown to vary depending on the position of the individual introducing the technology in the firm's organizational hierarchy. Others argue that the transfer process is to be based on the prevailing market forces and the bidding and contracting systems employed (*2*). Alternative proposed processes are based on an overall consideration of technical, economical, and risk assessment factors using decision monographs and flow charts (*3*), on cost-benefit analyses (*4*), or on pairwise comparisons (*5,10*).

The conceptual framework of the decision-structuring process proposed in this research is shown in Figure 2. The framework incorporates the decision determinants identified in the previous section under alternative approaches to decision making. It starts by studying construction projects at the operation level to select the operations most suitable as candidates for new technology. The selection is done with the help of a heuristic-based module that evaluates the candidacy of operations using the following criteria:

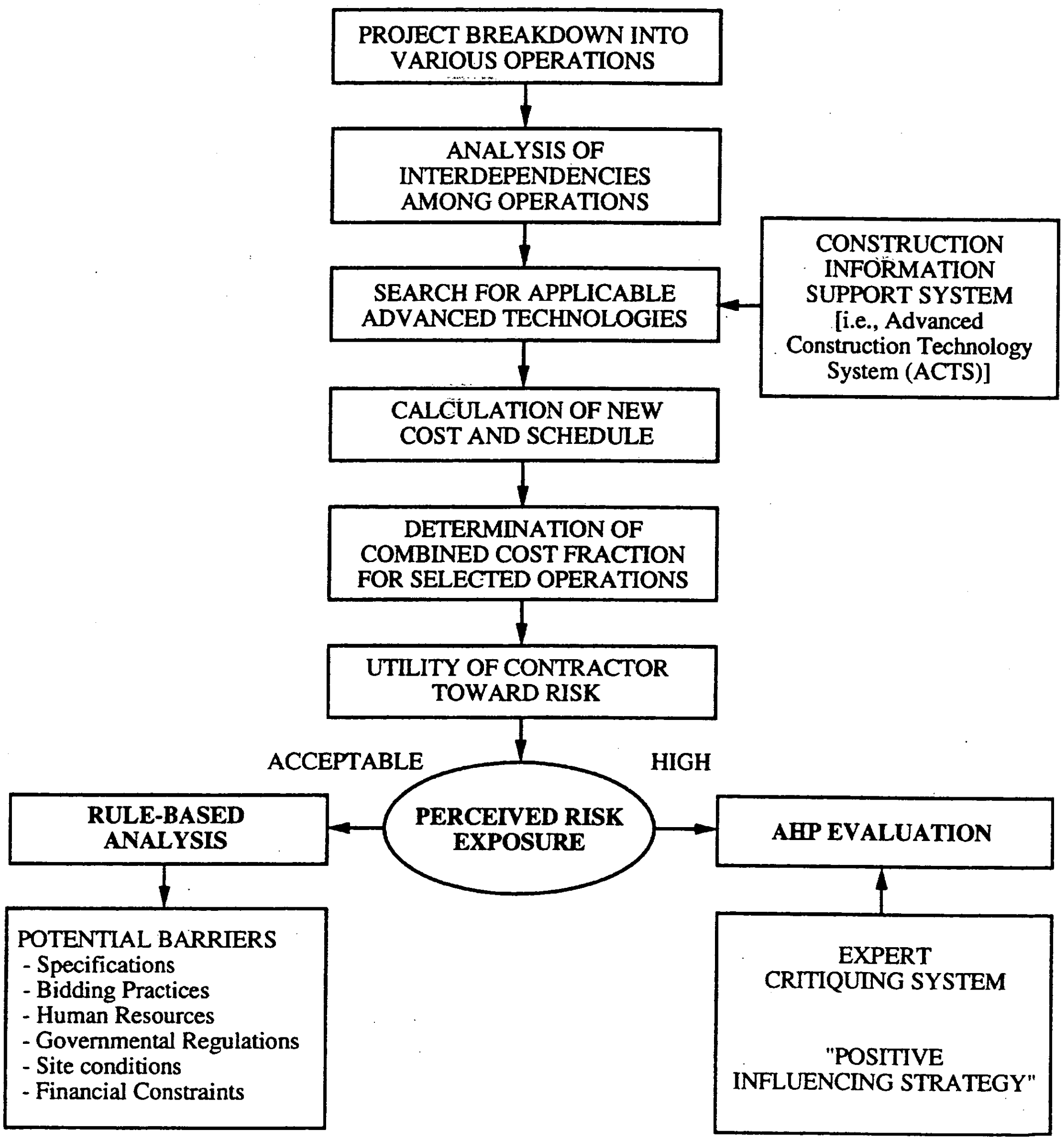


FIGURE 2 Conceptual framework for technology transfer and diffusion decisions.

1. Operation on critical or near-critical paths with a deterministic scheduling analysis, or operation with a high probability of being critical with probabilistic scheduling analyses;
2. Operation duration as a fraction of total project duration; and
3. Operation cost as a fraction of total project cost.

Candidate operations are then analyzed for their interrelationships with other operations. The list of candidate operations is expanded to include those related to the listed operations by start-to-start, finish-to-finish, and other forms of progress-based relationships. This step is particularly important because of the linear characteristic of highway construction work.

Next, a search for applicable advanced technologies is performed. It is proposed that a construction information support system such as the Advanced Construction Technology System (ACTS) be used to retrieve information documented on emerging construction technologies (*11*). The types of information that can be retrieved from ACTS include description, costs, benefits, limitations, experience, and operating environment, among other. The ACTS data base was developed at the University of Michigan with support from the Construction Industry Institute, which is taking steps to make it commercially available to the construction industry.

The new cost and schedule information based on the advanced technologies found to be applicable to selected candidate operations is used to determine a project's revised cost and schedule, which are incorporated in the decision-making process at later stages. To quantify the level of failure risk that contractors may assume by choosing to incorporate new technologies in the implementation of prospective projects and therefore decide on the level of vigorousness needed for the evaluation process, the combined cost fraction for the selected operations is determined (*4*). This figure is believed to represent the portion of the project's worth that contractors would be risking by applying new technologies; consequently, it is used along with a contractor's utility to judge which decision analysis approach would have to be chosen to satisfy the contractor's concern.

The utility of a contractor is approximated by examining a contractor's replies to a series of questions dealing with possible levels of loss of wealth. The observations are solicited using the probability equivalent method, and the utility examination is performed by determining the best fit from three families of mathematical functions: exponential, logarithmic, and polynomial. The three functions have different implications about the risk attitude of a contractor. Of a particular interest are the quadratic and fourth-order functions of the polynomial group, in which the risk aversion of decision makers increases as the level of wealth grows. Such behavior is thought to be not uncommon in an industry in which equipment dependency is intensive. Yet, even though the growing number of technological innovations may be rendering the existing technology obsolete, contractors may be reluctant to abandon con-

TABLE 1 Qualifiers and Typical Rule of Knowledge-Based Specifications Module

SPECIFICATIONS MODULE		
	Qualifier	**Qualifier Applicable Options**
1	The method used for specifying the execution requirements is	performance/descriptive/reference standard/proprietary
2	The descriptive requirements used are evaluated to be	• lenient or restrictive in a way that it allows the use of the new technology • restrictive in a way that it does not allow the use of the new technology
3	The reference standard is	up-to-date (not up-to-date) that it allows (does not allow) the use of the new technology
4	The proprietary specifications used are	closed/open
5	The closed proprietary spec does	allow (not allow) the use of the new technology
6	Alternates to the specified execution requirements are	named (not named) in the specifications
7	The alternates named do	incorporate (not incorporate) the new technology
8	The open proprietary specifications do	permit (not permit) the bidder to submit requests for substitutions
9	The open proprietary specifications do	control (not control) candidate substitutions by having to meet performance requirements
10	The new technology under consideration does	meet (not meet) the performance requirements prescribed by the open proprietary specifications

A Typical Specifications Module Rule
RULE NUMBER: 3 IF: The method used for specifying the execution requirements is descriptive and The descriptive requirements used are evaluated to be restrictive in a way that it does not allow the use of the new technology THEN: SPECIFICATIONS DO REPRESENT A BARRIER - Confidence = 10/10

ventional technologies that still have remaining physical lives. The loss associated with a contractor's 50 percent utile is compared with the combined cost fraction of operations incorporating new technologies. If the cost fraction is less than the 50 percent utile loss, the level of risk may be judged acceptable, and the rule-based analysis of potential barriers is activated to advise the contractor on technical and other types of obstacles to the technology-transfer decision. For a cost fraction higher than the 50 percent utile loss, the AHP approach is initiated wherein technology-thrust factors are weighed against technology-impediment factors with direct input and judgment received from the decision maker.

RULE-BASED ANALYSIS OF POTENTIAL BARRIERS

The rule-based module for the analysis of potential barriers was developed using EXSYS, a general-purpose expert system development shell. Six potential barriers were investigated as part of this analysis: specifications, bidding practices, human resources, governmental regulations, site conditions, and financial constraints. It is presumed that any or a combination of these factors could render the decision to implement a new technology technically unfeasible, even if the contractor accepts the associated risk.

Rule-based modules were developed that test each of the potential barriers considered, with the exception that the financial constraints module was designed as a recommendation to be displayed upon the request of the user. Each of the rule-based modules consists of a set of qualifiers that describe the factor in question and a number of rules generated using the named qualifiers. The qualifiers were identified from the extensive literature search performed as part of this research, and the rules were validated using the expertise of the authors. Rules were generated in a hierarchical format that would ensure the consideration of all technically and conceptually feasible combinations of qualifiers. The user is given access to additional information related to qualifier interpretations and expanded rule recommendations that can be retrieved using a special help command. The set of qualifiers and a typical rule for three of the five rule-based modules are described in Tables 1, 2, and 3. The confidence of a rule's recommendation is expressed as a fraction of 10. Any negative recommendation accompanied by full confidence (10/10) implies that implementation decisions are not feasible; the opposite is true for a full-confidence positive recommendation. For all recommendations with imperfect, assigned confidence, the user is advised on how to overcome those uncertain situations.

The governmental constraints module consists of rules derived from qualifiers dealing with the forms of government control on the import of new technologies, which may be high customs fees or a total ban on importation. The control could be to protect a labor-intensive economy or domestic equipment manufacturing. In addition, the site condition module is based on only two qualifiers dealing with site accessibility and geometric features.

Finally, the recommendation concerning the financial constraints factor emphasizes that contractors should be capable of objectively

TABLE 2 Qualifiers and Typical Rule of Knowledge-Based Bidding Practices Module

BIDDING PRACTICES MODULE		
	Qualifier	Qualifier Applicable Options
1	The cost per unit of work using the new technology is	higher(lower) than that using the conventional technology
2	The bidding evaluation system incorporates	• more than one parameter • only one parameter (cost)
3	The other parameter(s) incorporated in the bidding system is (are)	• schedule • quality • both schedule and quality
4	The schedule required to accomplish the specified work is	shorter (longer) using the new technology
5	The quality parameter is	important (not important) in relation to • the life-cycle costs of the facility • the aesthetic aspects of the facility • both
6	The quality obtained using the new technology is	better than (worse than) that obtained using the conventional technology

A Typical Bidding Practices Module Rule

IF: The cost per unit of work using the new technology is higher than that using the conventional technology
and The bidding evaluation system incorporates more than one parameter
and The other parameter(s) incorporated in the bidding system is (are) both schedule and quality
and The quality obtained using the new technology is better than that obtained using the conventional technology
and The quality parameter is important in relation to the life-cycle costs of the facility
and The schedule required to accomplish the specified work is shorter using the new technology
THEN: BIDDING PRACTICES DO NOT REPRESENT A BARRIER - Confidence = 9/10

TABLE 3 Qualifiers and Typical Rule of Knowledge-Based Human Resources Module

HUMAN RESOURCES MODULE		
Qualifier		Qualifier Applicable Options
1	The existing workforce at your firm is	able (not able) to operate the new technology
2	The existing workforce at your firm is	able (not able) to maintain the new technology
3	The existing workforce at your firm is	acquainted (not acquainted) with safety procedures in operating and maintaining the new technology
4	The basic educational knowledge and skills of your workforce are	enough (not enough) for them to learn how to safely operate and maintain the new technology
5	Incentive programs that help your workforce be motivated to adapt to the new technology do	exist (not exist) in your organization
6	Your firm is	equipped (not equipped) to handle the training of your workforce on operating and maintaining the new technology
7	Outside training centers in the country or abroad are	economically available (not available) to train your workforce on operating and maintaining the new technology
8	Local labor with needed knowledge and skills to operate the new technology are	available (not available) in the market
A Typical Human Resources Module Rule		
IF: The existing workforce at your firm is not able to operate the new technology or: The existing workforce at your firm is not able to maintain the new technology or: The existing workforce at your firm is not acquainted with safety procedures in operating and maintaining the new technology and The basic educational knowledge and skills of your workforce are enough for them to learn how to safely operate and maintain the new technology and Incentive programs that help your workforce be motivated to adapt to the new technology do exist in your organization and Your firm is equipped to handle training of workforce on operating and maintaining the new technology THEN: HUMAN RESOURCES DO NOT REPRESENT A BARRIER - Confidence = 9/10		

judging the effect of possible marginal increases in bid prices on their chances of winning contracts. It is recommended that contractors' judgment be dependent on (*a*) the probable increment in bid price that can be attributed to higher financing and bonding premiums, relative to the total bid price; (*b*) the uncertainty inherent in the total bid price that is attributed to the technical aspects of the project; and (*c*) the policy adopted for markup determination along with the level of competition and need for work.

AHP EVALUATION

Background and Structure of Hierarchy

The AHP is a methodology for solving complex problems that involves many criteria using the knowledge, expertise, and judgment of the decision maker. By applying this technique to the technology transfer and diffusion problem, highway contractors are provided with hierarchy (Figure 3) in which all the relevant factors are organized in a logical and systematic way from the goal to the factors and subfactors, and down to the alternatives of technology choice.

Expert Choice, an AHP-based decision analysis software, was used to conduct automated analyses of the designed hierarchy; the basic principles of AHP are covered in the literature (*5,12*). In the AHP evaluation procedure, contractors are asked to judge the elements of the hierarchy as to their relative importance with respect to a higher-level criterion or property. The judgments are made using pairwise comparisons on a 1-to-9 numerical scale or its verbal equivalent. The pairwise comparisons are then synthesized to rank the alternatives from which the choice is to made.

Example Evaluation Problem

To illustrate how this evaluation is performed, the problem of selecting between the laser-based grading technology and the conventional technology is analyzed, and the analysis results are sum-

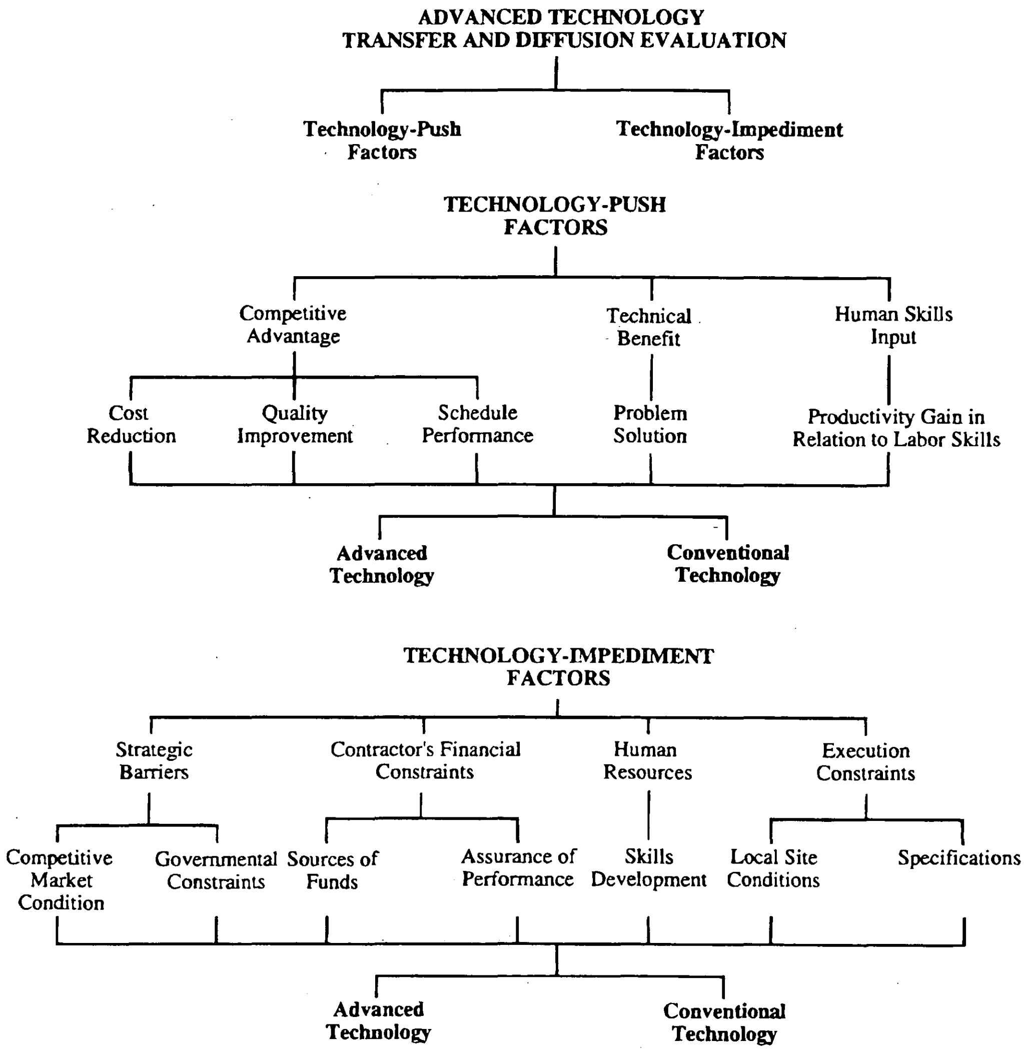


FIGURE 3 AHP for technology transfer and diffusion evaluation.

marized in Figures 4 through 6. Documented information on the technical properties and limitations of the advanced grading technology was assimilated from the literature and used, when appropriate, in the evaluation process (*1*). The abbreviations used in Figures 4 through 6 correspond to the elements of the hierarchy in Figure 3.

Three types of pairwise comparisons were used to provide judgments, examples of which are included in Figure 7. The term "importance" was used when comparing one criterion with another "preference" for comparing technology alternatives, and "likelihood" for comparing uncertain criterion occurrences. All comparisons are made with respect to higher-level criteria. Expert Choice tests the consistency of comparisons and helps the user improve it through an inconsistency measure.

In Figure 4 the AHP results synthesized at the factor and subfactor levels show the relative priorities of factors at the local (with respect to the next higher-level factor) and global (with respect to the goal) levels. For example, compared with the likelihood of being a barrier to the technology-transfer decision, the financial constraints factor, among the impediment, has a local priority of 0.649, which is higher than those of the other factors. This resulted from the comparisons given in Figure 7 with the financial constraints factor judged, with a value of 6, to be more important than all other factors in its group. In turn, the higher calculated priority indicates a greater contribution by this factor to the final decision. Similarly, the competitive advantage factor possesses the highest local priority, of 0.615, in the assessment of the thrusts for technology transfer.

The AHP results synthesized at the subfactor and alternative levels for the promotion and impediment subhierarchies are presented in Figures 5 and 6, respectively, with background information on typical comparison judgments also illustrated in Figure 7. As can be seen in Figure 6, the competitive market condition factor has a synthesized local priority of 0.833 compared with 0.167 for the government regulations factor. This higher priority is also attributed to the judgmental evaluation of the former factor to be strongly more likely (score of 5) to be a strategic barrier (relative to the next higher level). At the lowest level in the hierarchy the new technology is evaluated to be equally preferable to the old technology, as indicated by the judgment of 1.0 shown in Figure 7.

Figures 4, 5, and 6 refer not only to the local priorities calculated for the variables, but also to the global priorities that represent the

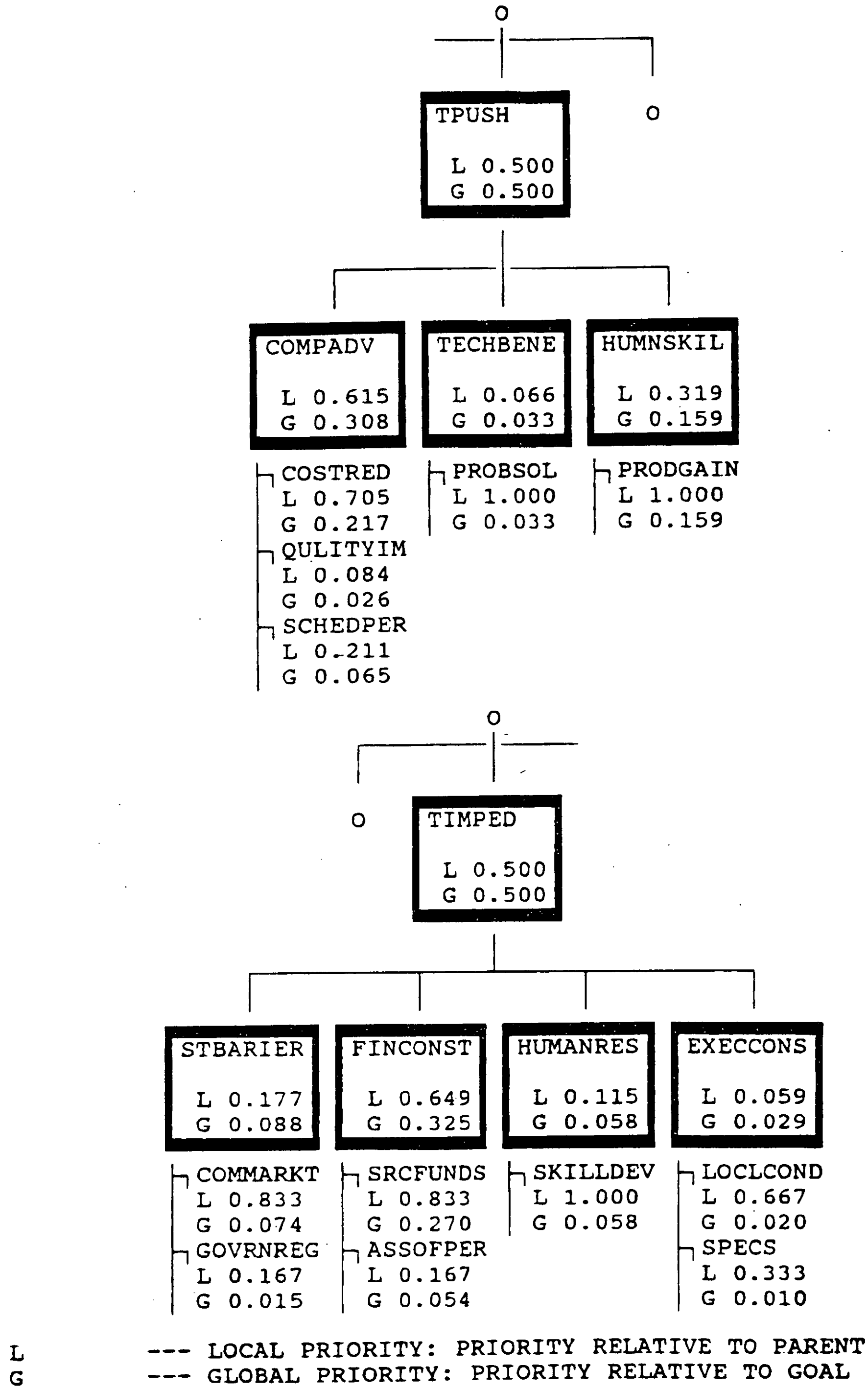


FIGURE 4 Synthesis of AHP evaluation results at factor and subfactor levels.

portion of the priority inherited by the various nodes. From the judgments used in this example, the synthesis of the evaluation with respect to the goal yielded a priority of 0.568 for the advanced technology, which compares with a priority of 0.4322 for the conventional technology, indicating that the former is slightly more preferred to the latter.

Sensitivity Analyses

Extensive analyses were performed to study the sensitivity of the decision to the input judgments used. The priorities of 0.568 and 0.432 generated at the goal level are based on equal weights given to both the negative and positive factors. The sensitivity of these

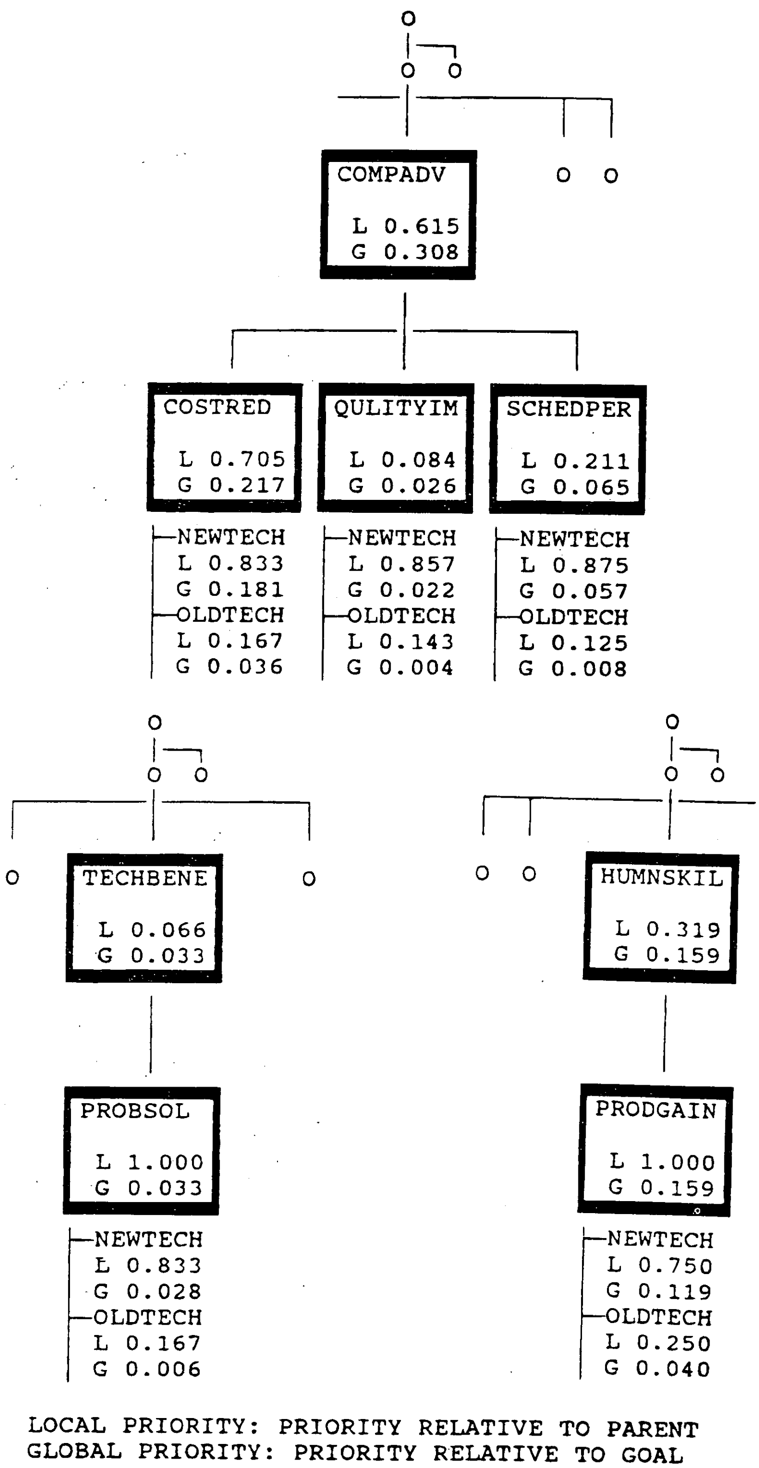


FIGURE 5 Synthesis of AHP evaluation results at technology promotion subfactor and alternative levels.

priorities to a change in the importance of the financial constraints factor is illustrated in the upper portion of Figure 8. The new technology becomes more preferable to the old one for lower calculated priorities of the financial factor, whereas the preference level decreases for higher priorities. However, the slopes of the goal priorities are not steep enough to intersect and, thus, induce a change of preference between the two choices (the new technology will always be preferred to the old one). In this dynamic analysis, when the priority level of the financial criterion is decreased or increased, the priorities of the remaining criteria increase or decrease proportionate to their original priorities, respectively. Under the impediment subhierarchy, the decision

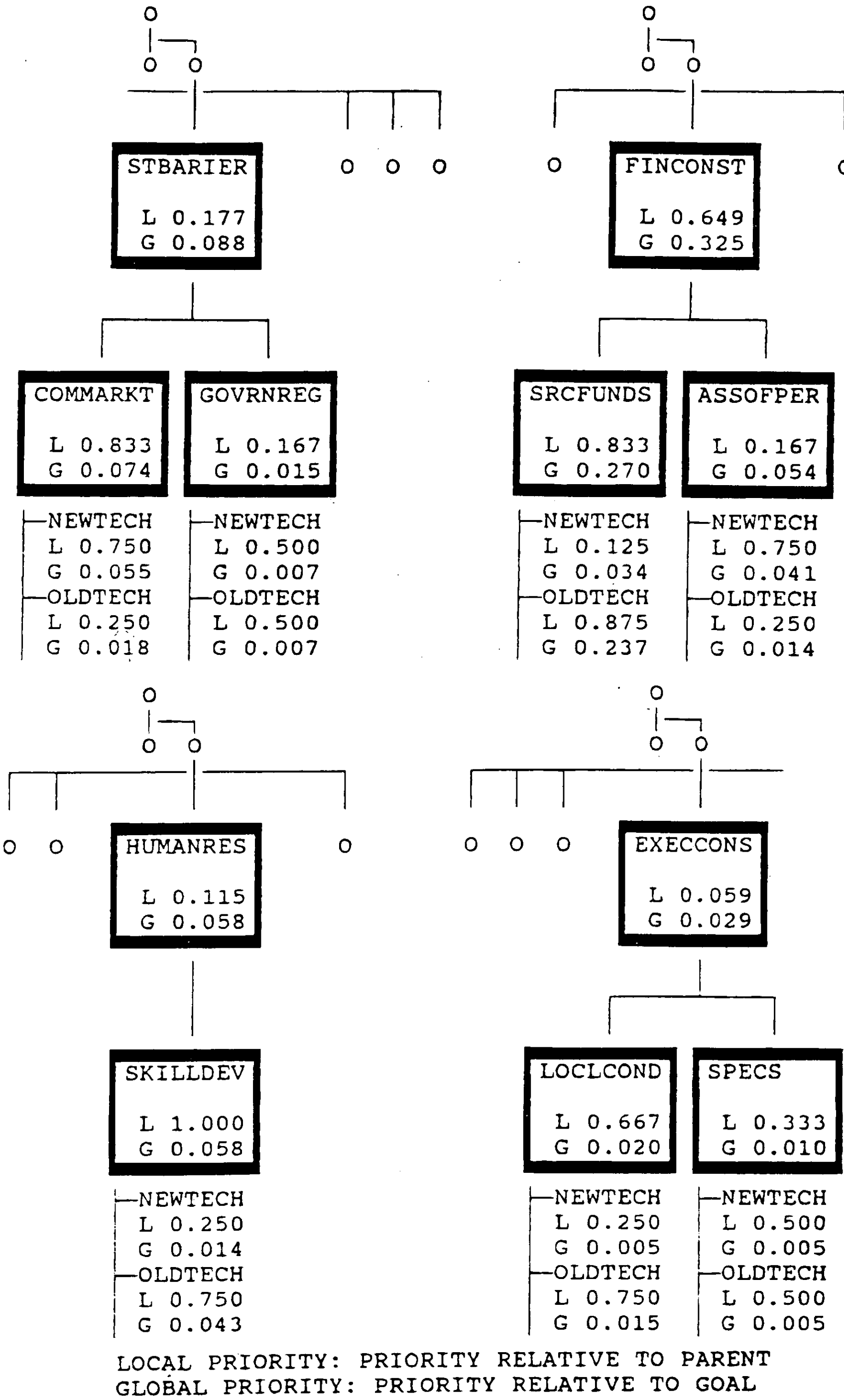


FIGURE 6 Synthesis of AHP evaluation results at technology-impediment subfactor and alternative levels.

was found to be similarly sensitive to the strategic barriers criterion and slightly sensitive to the human resources factor. However, almost no sensitivity was observed to the execution constraints criterion.

If the priorities are changed to 0.7 for the impediment node and 0.3 to the promotion node, the slopes of the goal priorities intersect as depicted in the lower portion of Figure 8. Here, the indication is that when the priority of the financial constraints factor is decreased to 0.487, the two technologies will be equally preferable. For higher priorities associated with the financial criterion, the decision will favor the old technology.

Expert Critiquing System

As discussed, the decision may be sensitive to the judgmental inputs used in quantifying the relative importance, likelihood, and preference of the identified criteria. Therefore, contractors considering the

JUDGMENTS WITH RESPECT TO
TIMPED < GOAL

	STBARIER	FINCONST	HUMANRES	EXECCONS
STBARIER		(6.0)	2.0	4.0
FINCONST			6.0	6.0
HUMANRES				3.0
EXECCONS				

Matrix entry indicates that ROW element is ___
1 EQUALLY 3 MODERATELY 5 STRONGLY 7 VERY STRONGLY 9 EXTREMELY
more IMPORTANT than COLUMN element unless enclosed in parentheses.

JUDGMENTS WITH RESPECT TO
STBARIER < TIMPED < GOAL

	COMMARKT	GOVRNREG
COMMARKT		5.0
GOVRNREG		

Matrix entry indicates that ROW element is ___
1 EQUALLY 3 MODERATELY 5 STRONGLY 7 VERY STRONGLY 9 EXTREMELY
more LIKELY than COLUMN element unless enclosed in parentheses.

JUDGMENTS WITH RESPECT TO
GOVRNREG < STBARIER < TIMPED < GOAL

	NEWTECH	OLDTECH
NEWTECH		1.0
OLDTECH		

Matrix entry indicates that ROW element is ___
1 EQUALLY 3 MODERATELY 5 STRONGLY 7 VERY STRONGLY 9 EXTREMELY
more PREFERABLE than COLUMN element unless enclosed in parentheses.

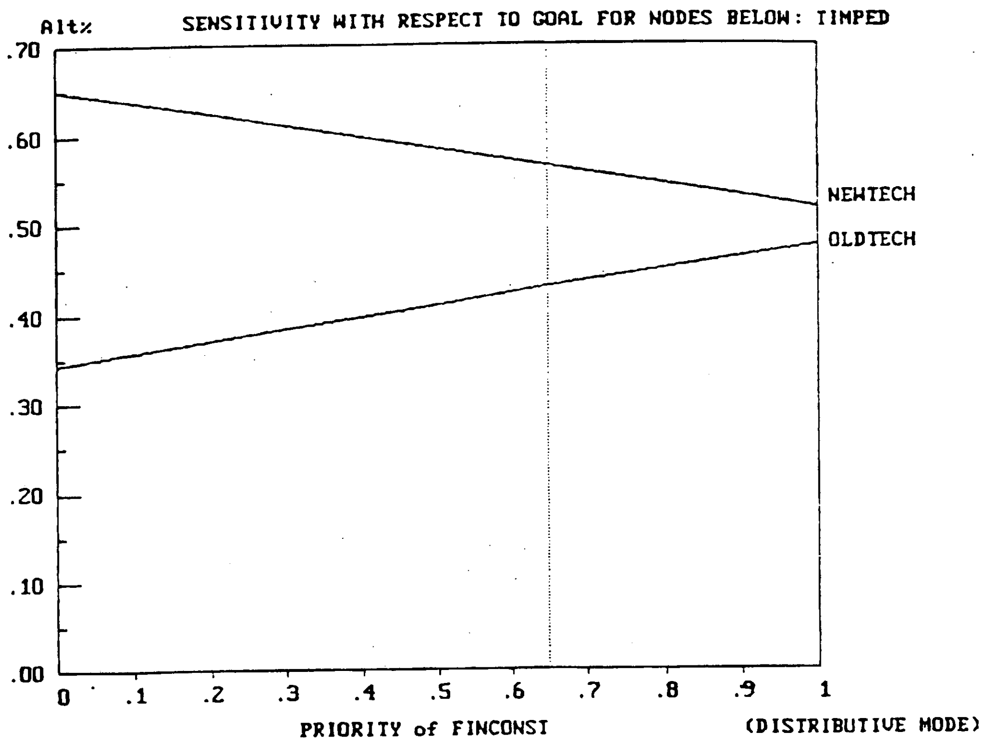
FIGURE 7 Three types of pairwise comparisons used in AHP evaluation.

feasibility of diffusing a new technology may have cognitive biases inherited in their intuitive judgments. Computer critics can be used to help overcome these biases. For the AHP evaluation, critiques can be made at all levels of the hierarchy. Namely, preference-based critiquing may be useful for weighing impediments against thrusts; likelihood-based critiquing may be employed for assessing factors and subfactors representing possible conditions and practices; and technical critiquing based on knowledge available on and experience gained with new technologies may be exercised to judge the preference of choices with respect to the various subfactors in the next higher level.

An expert critiquing system is under development that is intended to reduce the bias in the intuitive judgments used in the proposed hierarchical analysis. The critiquing system, which will be described in a future publication, with coverage of automation and interfacing properties of the support system, is designed to monitor the decision analysis process and counsel contractors on their reasoning and judgment in a way that positively influences the decision outcome (*13*).

SUMMARY AND CONCLUSIONS

A framework for structuring the process of technology transfer and diffusion decisions has been proposed. It incorporates a number of criteria identified to be significant to the decision-making process. The incorporated criteria are analyzed using either of two evaluation approaches that employ documented relevant construction information. Through an expert critiquing system, the information generated along the decision process—especially that of the rule-



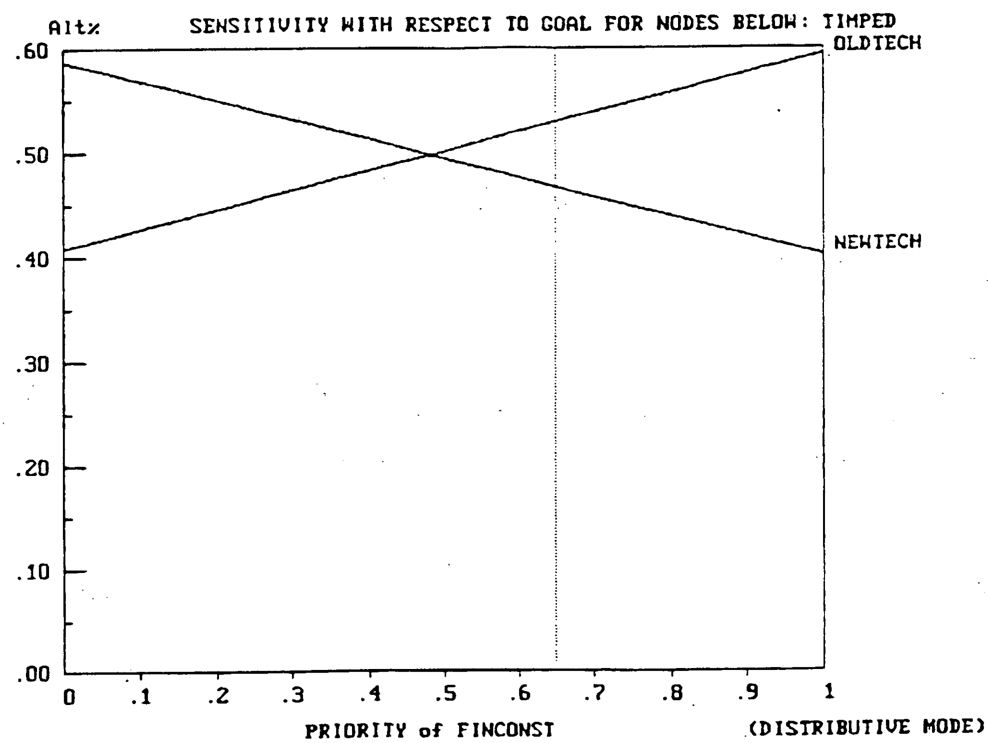


FIGURE 8 Dynamic sensitivity analyses showing interaction between hierarchy levels.

based analysis—can be made useful in the hierarchical evaluation approach to help remove possible bias from intuitive judgments.

ACKNOWLEDGMENTS

The financial support of this work by the University Research Board of the American University of Beirut is gratefully acknowledged. The technical support provided by the Engineering Management Department and facilities of the George Washington University is also appreciated. Finally, thanks are due to Anwar Karam, a former graduate student, and Mohammad Moukahal, a current graduate student, for their assistance throughout this research work.

REFERENCES

1. Tatum, C. B., and A. T. Funke. Partially Automated Grading: Construction Process Innovation. *Journal of Construction Engineering and Management,* ASCE, Vol. 114, No. 1, March 1988, pp. 19–35.
2. Johnson, K. D., and C. B. Tatum. Technology in Marine Construction Firms. *Journal of Construction Engineering and Management,* ASCE, Vol. 119, No. 1, March 1993, pp. 148–162.
3. Lutz, J. D., L.-M Chang, and T. R. Napier. Evaluation of New Building Technology. *Journal of Construction Engineering and Management,* ASCE, Vol. 116, No. 2, June 1990.
4. Chang, L.-M., D. E. Hancher, T. R. Napier, and R. G. Kapolnek. Methods to Identify and Assess New Building Technology. *Journal of Construction Engineering and Management,* ASCE, Vol. 114, No. 3, Sept. 1988.
5. Skibniewski, M. J., and L.-C. Chao. Evaluation of Advanced Construction Technology with AHP Method. *Journal of Construction Engineering and Management,* ASCE, Vol. 118, No. 3, Sept. 1992, pp. 577–593.
6. Miller, M. L., and L. E. Bernold. Sensor-Integrated Nailing for Building Construction. *Journal of Construction Engineering and Management,* ASCE, Vol. 117, No. 2, June 1991, pp. 213–225.
7. Tatum, C. B. Potential Mechanism for Construction Innovation. *Journal of Construction Engineering and Management,* ASCE, Vol. 112, No. 2, June 1986, pp. 178–191.
8. Cushman, N. S., C. H. Nam, and C. B. Tatum. Technology Transfer in Building Construction—Case of Seismic Design. *Journal of Construction Engineering and Management,* ASCE, Vol. 118, No. 1, March 1992, pp. 129–141.
9. De La Garza, J. M., and P. Mitropoulos. Technology-Transfer (T^2) Model for Expert Systems. *Journal of Construction Engineering and Management,* ASCE, Vol. 117, No. 4, Dec. 1991, pp. 736–755.
10. AbouRizk, S. M., S. R. Mandalapu, and M. Skibniewski. Analysis and Evaluation of Alternative Technologies. *Journal of Management in Engineering,* ASCE, Vol. 10, No. 3, May–June 1994, pp. 65–71.
11. Ioannou, P. G., and L. Y. Liu. Advanced Construction Technology System—ACTS. *Journal of Construction Engineering and Management,* ASCE, Vol. 119, No. 2, June 1993, pp. 288–306.
12. Saaty, T. L. *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation.* McGraw-Hill International Book Co., London, England, 1980.
13. Silverman, B. G. *Critiquing Human Error: A Knowledge Based Human-Computer Interaction Approach.* Academic Press, London, England, 1992.

Publication of this paper sponsored by Committee on Construction Management.

BFX: Operational Expert System for Bridge Fabrication

W. M. KIM RODDIS, HANI MELHEM, MICHAEL R. HESS, AND SRINATH NAGARAJA

The Bridge Fabrication Error Solution Expert System (BFX) was developed to help designers and inspectors determine the severity of fabrication errors on steel bridge members and specify the necessary repairs. Aspects of the development, delivery, and operation of BFX of direct interest to highway bridge and materials engineers are described. The scope of BFX focused on tolerance, drilling and punching, cutting, and lamination fabrication errors that do not have a codified repair procedure. During predelivery testing, BFX provided the correct repair in two-thirds of the test cases, recognized that the test case was not covered by its rule base in one-third of the test cases, and gave the wrong solution for none of the test cases. BFX has been in use at the Kansas Department of Transportation since January 1994. An operational example using BFX is presented.

Errors arising during the steel fabrication stage may have a catastrophic effect on the performance of a completed highway bridge. More commonly, fabrication errors can cause delays in the fabrication process. All the information needed to support a good decision may not be available at the right time and in the right place to solve the problem in the restricted time necessary to keep the job on schedule. The Bridge Fabrication Error Solution Expert System (BFX) was developed to help design engineers and materials inspectors determine the extent of damage due to fabrication errors and specify the necessary repairs. In addition, BFX is intended to be used as a training tool for novice bridge engineers and material inspectors.

BFX was created to provide a unified repair procedure for the Kansas Department of Transportation (KDOT) by gathering domain expertise from designers, inspectors, and fabricators. The goal was to create a system that would provide the most suitable repair solution in the most timely manner. Within this context, it was decided that no answer was preferred to a wrong answer. A design objective was thus a system that would indicate clearly when a submitted problem was beyond the system's scope. The system focuses on fabrication errors that do not have standard code specifications for repair. The completed expert system was delivered to KDOT in January 1994 (*1*).

DEVELOPMENT

The project used expert system software tools and development methodologies tailored specifically to KDOT's mission and needs. The development strategy was designed to deliver a system that would address the real needs of KDOT and would become a functional tool for determining solutions to fabrication errors. Details on the different development stages and the approaches used can be found elsewhere (*2*). The success of expert system development projects is also highly dependent on establishing interaction with target users at an early stage of the project and maintaining this contact throughout the development cycle. To meet these requirements, the system designers used a panel of experts and a panel of users; each panel consisted of six individuals, including design engineers, material inspectors, and a fabricator. The panel of experts was created to resolve steel fabrication error solutions. Panel members gave their expertise on fabrication processes and procedures and acted as the primary experts for knowledge acquisition. The panel of users was established to target users of the developed system. This panel provided information for the scope of the system and interface design.

Each panel had representative members from design, inspection, and fabrication. Gathering experts from all three areas involved with bridge fabrication—design, inspection, and fabrication—allowed more interaction and provided broader information on conditions of errors and repair solutions. Experts in each of the individual areas are exposed to particular parts of a fabrication error. By using representatives of these areas in panel meetings and other interviews, the development system more accurately provided detailed solutions and conditions for fabrication errors. When the panels were formed, the members understood that they would be required to participate in panel meetings and personal interviews, provide data cases, respond to questionnaires, and review the system. The total time spent by all panel members combined was between 2 and 3 person months. This time includes panel meetings, collection of cases, knowledge acquisition interviews, evaluation of system, and training. Table 1 presents the estimated time commitment for panel participants.

SCOPE

BFX deals specifically with errors due to tolerance (dimensional), drilling and punching, cutting, and lamination. The scope of the system was developed using modules and submodules, organized as shown in Figure 1. The tolerance portion of the scope contains mislocated holes, edge distance, end distance, mislocated member, miscut member, misattached member, misaligned member, and stress fracture submodules. The drilling and punching portion contains procedures, misshapen holes, partially drilled holes, and size submodules. The cutting section contains nicks and gouges, mismilled edges, and miscut orientation submodules. Lamination contains surface, internal, and edge submodules. It was very important to limit the scope during system development so that the design criteria could be applied effectively and in more detail.

W. M. K. Roddis, Department of Civil Engineering, University of Kansas, 2008 Learned Hall, Lawrence, Kan. 66045. H. Melhem and S. Nagaraja, Department of Civil Engineering, Kansas State University, 119 Seaton Hall, Manhattan, Kan. 66509. M.R. Hess, HNTB Corporation, 7007 College Boulevard, Suite 150, Overland Park, Kan. 66211.

TABLE 1 Estimated Time Commitments for Panel Participants

Type of Participation	Panel Participation: Number	Panel Participation: Length (hours)
Panel	4	4
Personal Interview	4	2
Questionnaires	2	4
Panel Meetings	2	4
System Review	3	2

The system was developed and delivered using the Level5 Object shell (*3*), chosen as a standard for KDOT, running on PC-486 machines. The design of the system and the use of Level5 Object development tool allow knowledge to be added and the system to be modified. Many failures with expert systems can be attributed to creating too broad a scope. Success is more likely if goals are well defined and allowance is made for the addition of new knowledge or as other areas of need are defined. BFX was developed with these principles in mind.

There is a broad range of severity that can occur from fabrication errors. The degree of severity depends on the type of error and on the member in which the error occurs. Depending on their severity, many errors are handled entirely within a fabricator's own shop inspection system. Some errors require contact with state inspectors or design engineers for an approved repair method. Some types of fabrication error have standard solutions but still require contact with state inspectors for approval. The knowledge-based expert system provides a "best" solution and any other allowable solutions. The system also documents the basis of the repair solution and, if requested, a history of the knowledge path.

KNOWLEDGE ACQUISITION AND ENGINEERING

The development of an expert system requires both knowledge acquisition and knowledge engineering. The knowledge acquisition stage consists of gathering the knowledge from experts. The knowledge that is elicited or acquired from the expert sources is used to build the knowledge base. The knowledge engineering stage consists of translation and transformation of problem-solving expertise from a knowledge source to a human or computer program destination. Knowledge engineering is thus the process of mapping the knowledge gathered from experts into a programmed knowledge base.

The development, verification, and validation of BFX all depended on the availability of many accurate example cases and interaction with panel members. The knowledge acquisition methodology chosen therefore focused on collecting actual cases of past fabrication errors and successful repairs. Gathering information to put into BFX occurred in different stages. The first step was to gather case examples directly from fabrication shops, state inspectors' field notes, and bridge project documents. Next, individual interviews were conducted using case studies and hypothetical data case examples based on variations of the actual data cases gathered and interview sessions. Using actual and hypothetical cases, the solution sets for multiple types of errors were determined. Finally, the repair solutions generated were approved by design engineers and inspectors and verified by certified design procedures. Using the information gathered, rules were developed and

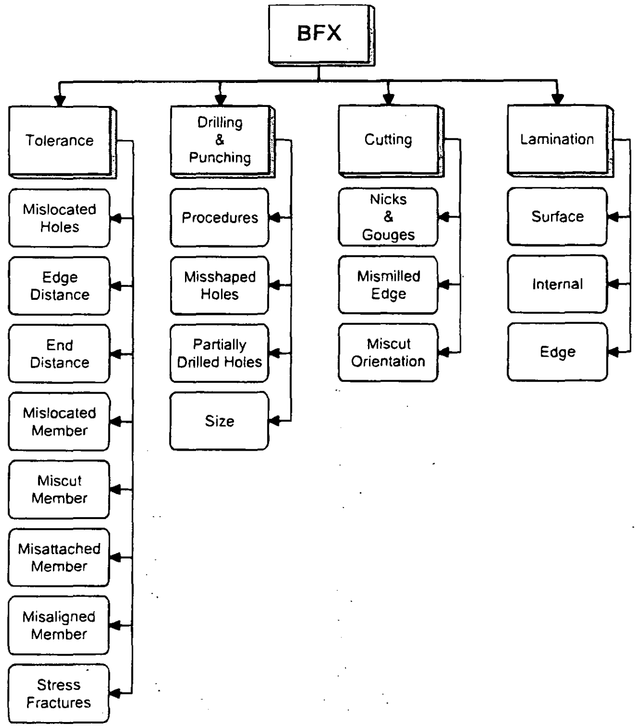


FIGURE 1 BFX knowledge modules and submodules.

implemented into the Level5 Object shell. An automatic generation of rules from the case examples was also investigated using inductive techniques established by artificial intelligence researchers in computer science. The assessment of this investigation is discussed elsewhere (*4*).

These data cases were checked further against technical specifications and documentation of current procedures. These case examples were collected from experts' questionnaires to KDOT bridge engineers, fabrication personnel, and inspectors; historical records such as case studies, maintenance data bases, and inspection reports; and simulation results that were generated internally. Actual data cases were cataloged and checked for completeness; from these actual data cases, hypothetical data cases were created by the knowledge acquisition team to be used during individual interview sessions. The collection of actual cases was partitioned into development examples to be used for knowledge acquisition and test cases to be used for validation and verification. The distribution of the 77 actual cases used for development is shown in Figure 2. The percentage distribution of the development cases may be assumed to give a rough measure of the distribution of error types encountered in practice by KDOT, since the development cases were collected from past KDOT experience.

The personal interviews included one-on-one sessions and, in some cases, two panel members per interview session. These interview sessions were used to gather specific information about certain data cases provided by panel members and to answer hypothetical variations of these data cases. In addition, these sessions were used to discuss the rationale of certain repair solutions associated with problem types described in the data cases. These data cases provided by panel members were actual errors that had occurred during fabrication and were resolved at the fabrication shop. The cases described the errors and their repair solutions.

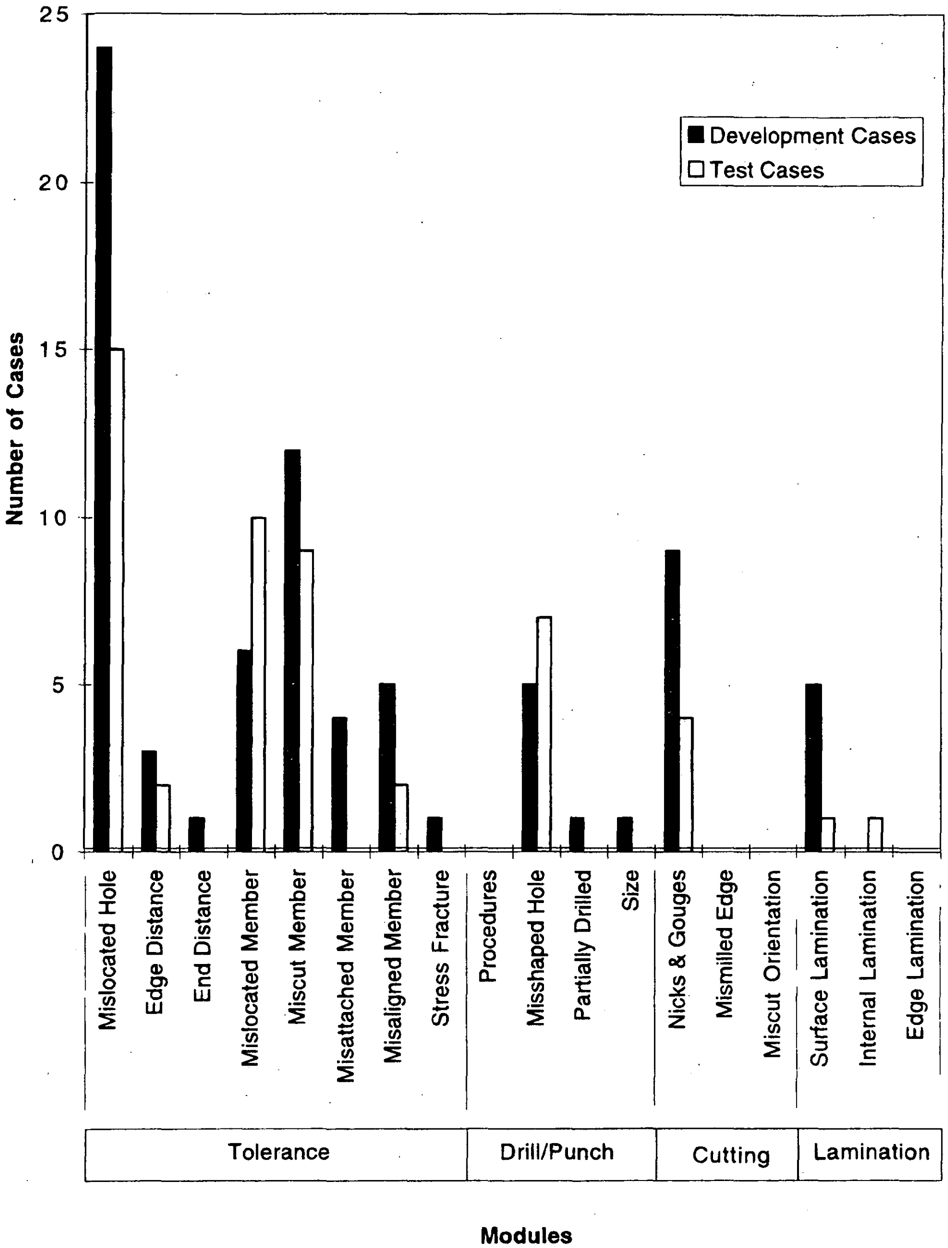


FIGURE 2 Distribution of development and test cases.

More data from the interviews were gained by structuring the interviews around developing repair solutions for prepared actual and hypothetical cases. Information from actual data cases was also verified by panel members during the interview sessions. Secondary interviews were used to finalize the clarification of hypothetical data cases and information on technical specification requirements. Interview sessions began by covering actual data cases and clarifying any incomplete information needed for specific data cases. Hypothetical data cases were then presented, and repair solutions completed with corresponding information. The documented actual data cases were modified to be hypothetical to collect more information and get as complete a coverage of error cases as possible. The hypothetical cases were used to address issues arising from the knowledge base development. The documented data cases were also reviewed during the interviews for confirmation on the repair procedures given. The hypothetical cases included minor and major changes in actual data cases. Repair solutions given for these hypothetical data cases were checked by presenting the cases at subsequent interview sessions with other panel members. Once completed, the cases were included in the prototype development system. Data cases were then transformed into rules for the system program and assisted the design team in understanding the experts' problem-solving techniques.

PERFORMANCE

The capabilities of the BFX were checked by testing the system. Validation and verification of the system were based on two methods. The first method was the actual running of the system using 18 hypothetical test cases by the expert and user panel. The second method was a performance check of the system using 33 test cases that were not used in the development of the system and that met the scope of the system.

The first method of the validation and verification testing was completed on the pilot delivery program using the panel members. The system was then evaluated by using hypothetical test cases provided by the members. Realistic conditions were simulated by having the panel members perform the input and run the cases by themselves. The hypothetical cases were based on actual problems experienced by experts. The total 18 panel test cases resulted in 11 correct solutions, 6 no solutions, and 1 incomplete solution. This first form of testing thus resulted in 61 percent correct solutions, 33 percent no solutions, and 6 percent incomplete solutions. When a fabrication error case is run on the system and no match between that particular type of error and the knowledge base occurs, the system will inform the user and suggest that the error case be implemented into the system. No match between the test cases and the knowledge base occurs when these types of fabrication errors have not been found during development of the knowledge base.

A program use questionnaire was provided to panel members after each validation and testing session. The panel members were asked to grade the performance of the system on the basis of the validation and verification criteria and relate information on performance and interface use of the system. Figure 3 presents the results for each module; a score of 5 is very good and a score of 1 is poor. Performance was graded as average to slightly above average in fulfilling expectations of the depth and accuracy of the system. The testing members were very impressed with the overall development and performance of the system based on initial projections.

The second method of validation and verification testing was checking the performance of BFX using 33 actual cases provided by panel members. These cases had not been used in system development and met the scope of the system. After running the 33 test

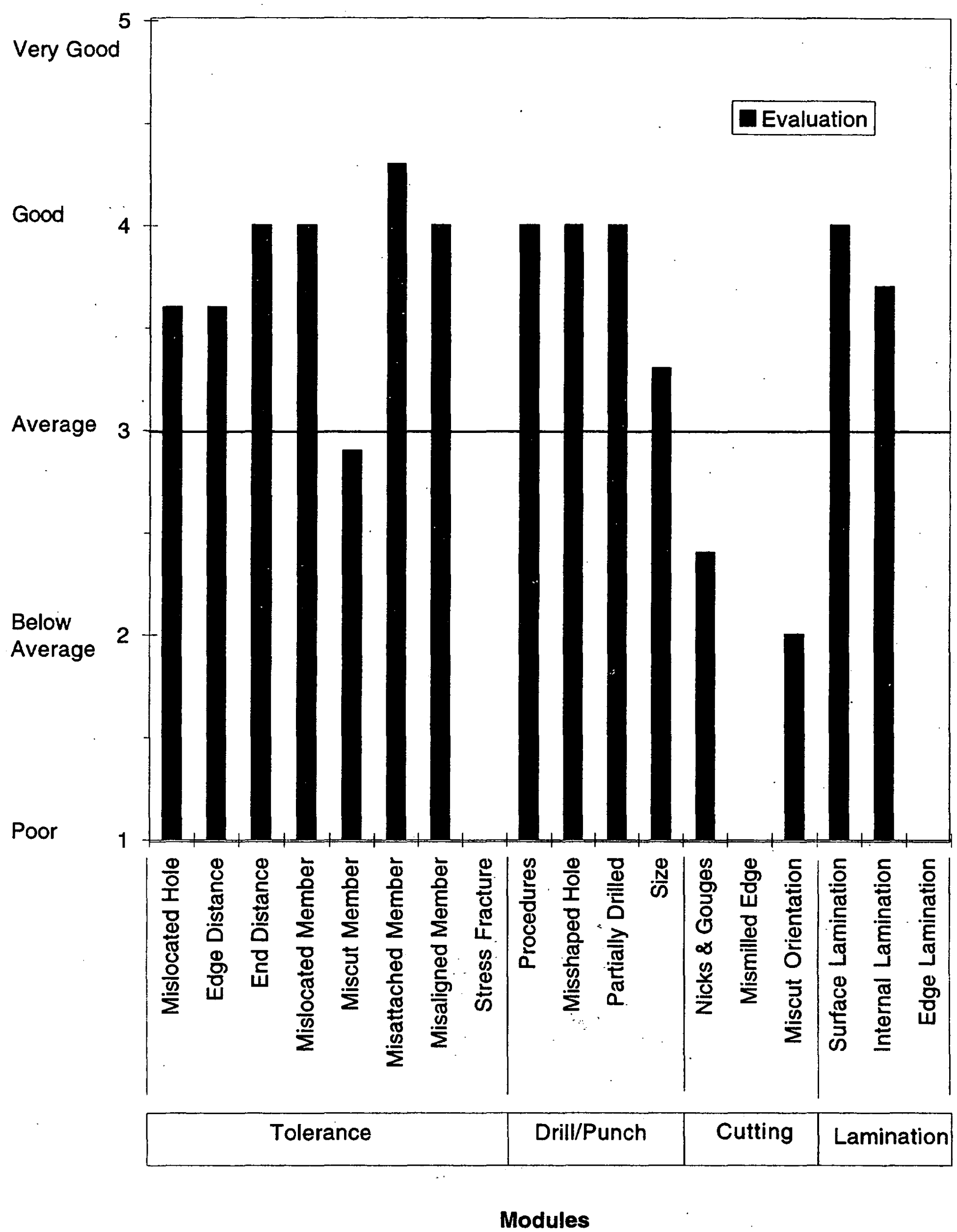


FIGURE 3 Panel evaluation of submodule performance.

cases, 21 of the cases gave the correct repair solution for each case. Twelve of the cases did not match the contents of the knowledge base during runs of the system. This second form of testing thus resulted in 64 percent correct solutions and 36 percent no solutions.

No logic errors occurred during any testing stage of the system, which shows that in terms of reliability, the system performed very accurately. This is very important in building user confidence; it is much better to receive no answer than an incorrect one. Combining the 18 hypothetical panel test cases and the 33 actual test cases resulted in 51 test cases distributed as shown in Figure 2. The distribution of development cases by module roughly matches the distribution of test cases by module. Since the development cases were collected from past KDOT experience, it may be inferred that the distribution of development cases by module also roughly matches the distribution of error types encountered in practice by KDOT. Combining both validation methods, BFX reached the correct solution in 63 percent of the cases, determined that the case did not match the contents of the knowledge base and therefore did not make a recommendation in 35 percent of the cases, and provided an insufficiently detailed recommendation in 2 percent of the cases. These results are shown in Figure 4. The BFX system and its performance results were presented to the bridge community and were well received (*5*).

DELIVERY

A successful expert system is one that can be maintained and kept current to accommodate new fabrication errors introduced to the system. To address the issues of maintenance and modification, a training seminar was established on BFX for KDOT personnel. This 2-day training seminar was presented at the offices of KDOT. Its purpose was to familiarize KDOT personnel with the technical specifications of the knowledge base and provide sufficient instruction for them to perform basic maintenance on the program without outside assistance. A training manual was prepared to be used during the seminar and in future training (*6*). Training consisted of an introduction to pertinent features of the development tool Level5 Object, an overview of the construction of the knowledge base for the fabrication expert system, and hands-on examples describing how to modify the knowledge base for basic maintenance.

BFX has been in use at KDOT since January 1994. A user's manual was prepared to help new users run the system and understand how it works (*7*). One of KDOT's bridge engineers has made two implementation changes during that time. When originally delivered, BFX did not provide an output file echoing each session input and results. The "session history" feature on Level5 Object, expected to fill the need for such a record file, was found to be awkward and unsatisfactory in practice. BFX was modified to echo input and output to a text file for complete and easy documentation of all cases. A second change deals with the input screens. Many of the input screens request multiple pieces of information on a single screen. Originally, the same input screens were used in various submodules, leading to cases in which some of the requested input was not used. These superfluous input requests have been deleted.

Tracking the system during the first year of operation is important. BFX is being used and tested, with each case being documented for performance and user comments. It is necessary that any areas of incomplete coverage within BFX be documented and revised to meet user needs. The tracking of BFX during initial stages of modification also shows the accuracy of the solutions and coverage of the knowledge domain. Documenting the performance of the system enables it to be modified when necessary. Documentation of the performance includes copies of the error and information on how BFX performed during operation of the error case. This information allows KDOT to make changes to BFX as necessary and to add missing error types to the knowledge domain. Documentation of the performance also allows KDOT to judge the accuracy of BFX and increase confidence in the repair recommendation given by the program. KDOT bridge personnel are maintaining and expanding the knowledge base.

Few operational cases have been run to date on BFX. On the basis of the small sample available, some preliminary findings can be stated. First, coverage of fabrication error types is incomplete, with at least a quarter of the cases resulting in no solution. This incompleteness of the knowledge base was expected because only a few fabrication error types were covered by the development cases. BFX was designed to allow the addition of knowledge to the system and increases in its scope. The system was segmented into individual submodules to allow easier modification and maintenance, with each submodule corresponding to an individual scope area of the system. Initial indications are that this approach was successful and that adding error types to the knowledge base is straightforward.

Second, the cutting module and its constituent submodules (nicks and gouges, mismilled edge, and miscut orientation) require refinement. On one operational case involving a sawcut gouge partially through a coverplate on a rolled beam, BFX recommended a weld repair when a grinding repair was more appropriate. On another operational case involving a web gouge due to a flame-cutter mis-

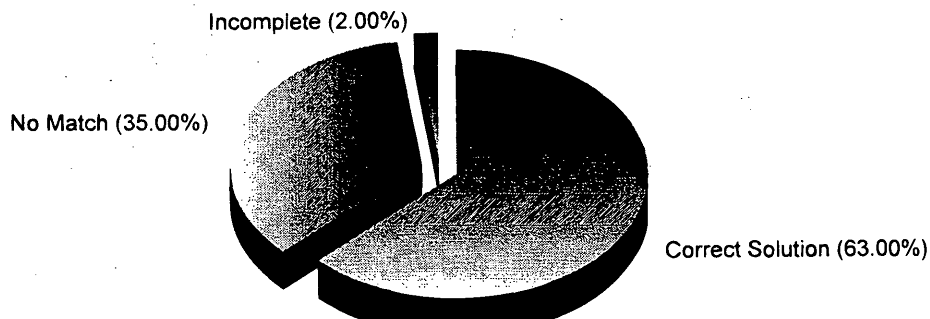


FIGURE 4 Test results.

tracking, BFX recommended replacement of the member when patching with an inset web slug was more appropriate.

One operational case involving several uses of BFX is presented to demonstrate BFX's capabilities. This example deals with mislocated holes at a plate girder flange splice. Several holes were misdrilled in the bottom flange of a plate girder, as shown in Figure 5.

The hole mislocations resulted in a variety of fabrication errors. First, the specified splice plate will no longer fit the hole locations in the bottom flange. This problem was entered into the tolerance module of BFX with the mislocated holes submodule selected. The input described the lack of fit problem. BFX's recommended solution was to leave the hole in the main member and make a new

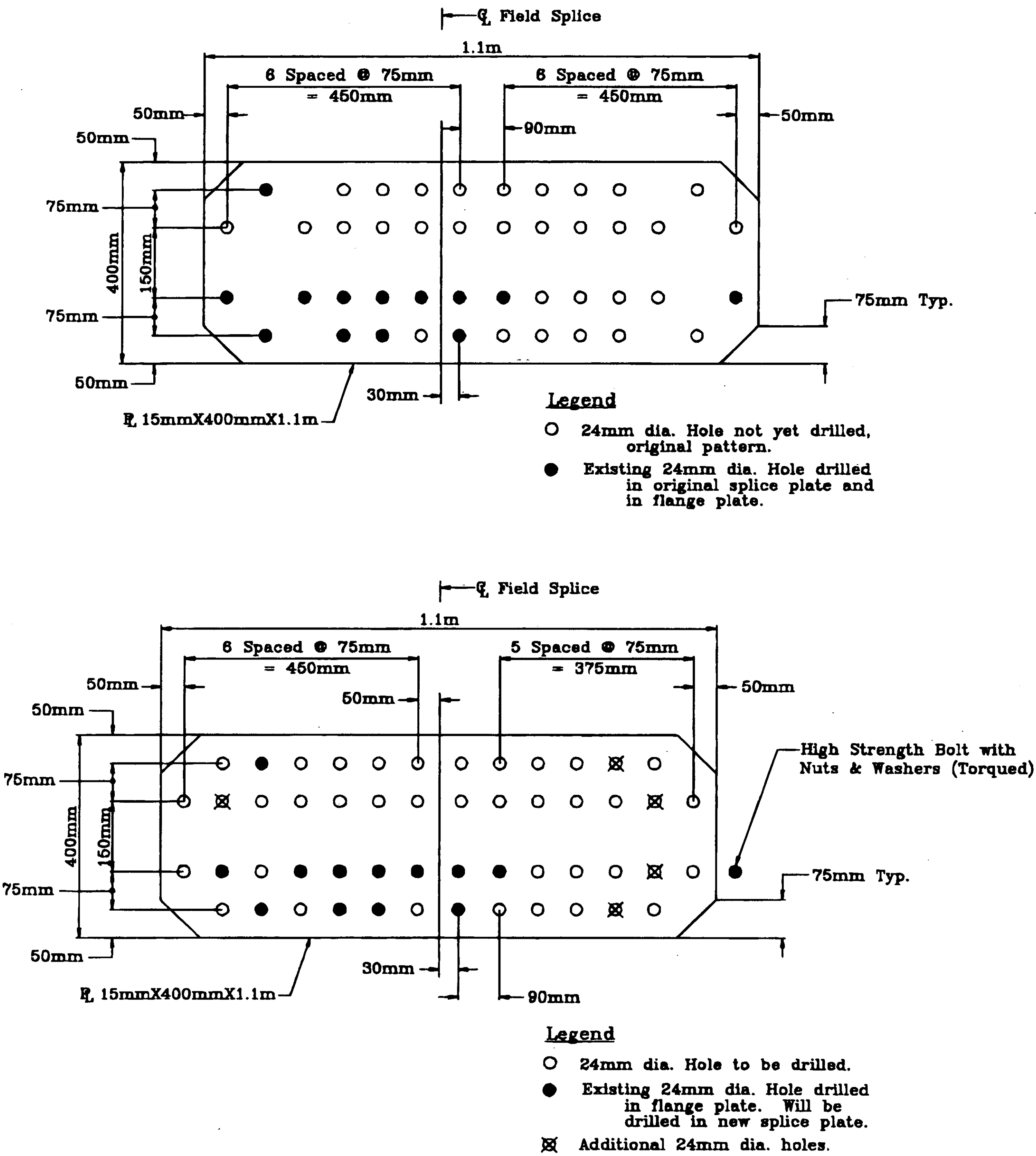


FIGURE 5 **Operational example: mislocated holes at flange splice:** ***top,*** **misdrilled,** ***bottom,*** **modified (1 in. = 25 mm).**

splice plate to match the existing hole pattern. The repair specified in Figure 5 does indeed use this approach.

Second, the mislocated hole on the extreme right is superfluous since it begins an additional row beyond those specified. This problem was entered into the tolerance module of BFX with the mislocated hole submodule again selected. The input this time described the extra bolt line problem. BFX's recommended solution was to leave the splice in the specified location and then take one of the following options: (*a*) extend the splice plate to cover the mislocated holes and drill to match, or (*b*) place bolts and washers in the additional holes and leave the splice plate as designed. The repair specified in Figure 5 takes the second approach.

Third, the two mislocated holes immediately to the right of the splice centerline violate end distance requirements. This problem was entered into the tolerance module of BFX with the end distance submodule selected. BFX's recommended solution was to add additional bolts in the bolt line if possible or cut and replace the member if not possible. The repair specified in Figure 5 takes the first approach. The total repair specified in Figure 5 thus is a superposition of the three approaches recommended by BFX for the three individual problems generated by the hole mislocations.

CONCLUSIONS

The development, delivery, and initial use of BFX has resulted in the following conclusions:

- BFX achieved the performance expectations desired by KDOT.
- BFX achieved the desired scope and accuracy established by KDOT. The knowledge domain was very suitable for development.
- The development methodology of using panels of experts was successful for this project.
- The modular development of BFX was easier and will simplify maintenance and modifications by KDOT.
- BFX is making a successful transition from an academic development environment to an operational system. As anticipated, the knowledge base must be expanded to cover fabrication error types not included in the development cases.
- The training provided enables KDOT bridge engineers to maintain and expand BFX, allowing KDOT to maintain and update the system.

The system has shown that it provides consistent, logical solutions to the fabrication errors specified in the scope of the development. As users of the system become more confident in the system's repair recommendations, fewer checks with design engineers will be necessary. Increases in the repair turnaround time will be achieved, which will reduce costs for fabricators and ultimately for KDOT. BFX can also be used to help train new inspectors and designers in the repair of fabrication errors. Inspectors and designer will also become more familiar with recognizing possible problems. The use of BFX will also provide better documentation and record keeping for KDOT. BFX is proving to be a useful system for KDOT, and consideration is being given to generalizing and expanding BFX so that it could be used by other states or regions.

ACKNOWLEDGMENT

This work was funded by the Kansas Department of Transportation Cooperative Research Program, an ongoing, cooperative, and comprehensive research program addressing transportation needs for the state of Kansas.

REFERENCES

1. Hess, M. R, W. M. K. Roddis, S. Nagaraja, H. G. Melhem, and A. Moran. *An Expert System for Fabrication Error Solutions.* Report K-TRAN: KU/KSU-93-2. Kansas Department of Transportation, Topeka, Aug. 1994.
2. Roddis, W. M. K., and M. R. Hess. Bridge Fabrication Error Solution: Developing a Knowledge-Based Expert System for a State Department of Transportation. *Proc., 1st Congress on Computing in Civil Engineering* (K. Khozeimeh, ed.), ASCE, New York, 1994, pp. 1687–1694.
3. *Level5 Object: Object-Oriented Expert Systems for Microsoft Windows.* Information Builders, Inc., New York, 1993.
4. Nagaraja, S., and H. Melhem. An Inductive Knowledge-Based System for Mitigation of Bridge Fabrication Errors. *Proc., 1st Congress on Computing in Civil Engineering* (K. Khozeimeh, ed.), ASCE, New York, 1994, pp. 1083–1090.
5. Melhem, H., M. R. Hess, and W. M. K. Roddis. Bridge Fabrication Error Solution: Developing an Expert System for a State DOT. *Proc., 11th Annual International Bridge Conference,* Engineers' Society of Western Pennsylvania, Pittsburgh, 1994, pp. 416–423.
6. Hess, M. R, W. M. K. Roddis, H. Melhem, and A. Moran. *Bridge Fabrication Error Solution Expert System Training Manual.* SL Report 94-2. University of Kansas Center for Research, Inc., Lawrence, 1994.
7. Hess, M. R., and W. M. K. Roddis. *Bridge Fabrication Error Solution Expert System User's Manual.* SL Report 94-1. University of Kansas Center for Research, Inc., Lawrence, 1994.

Publication of this paper sponsored by Committee on Fabrication and Inspection of Metal Structures.