1156

TRANSPORTATION RESEARCH RECORD

Transportation Systems Planning and Management

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1988

Transportation Research Record 1156
Price: $15.50
Editor: Elizabeth W. Kaplan
Production: Lucinda Reeder

modes
1 highway transportation
2 public transit
5 waterways

subject areas
11 administration
12 planning
15 socioeconomics
17 energy and environment

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Transportation systems planning and management.
p. cm.—(Transportation research record, ISSN 0361–1981 ; 1156)
Includes bibliographies.
ISBN 0-309-04666-1
1. Transportation and state. 2. Transportation—Planning.
I. National Research Council (U.S.). Transportation Research Board.
II. Series.
TE7.H5 no. 1156
[HE193]
380.5 s--dc19
[380.5'068]
88-25126
CIP

Sponsorship of Transportation Research Record 1156

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION

Management and Finance Section

Committee on Manpower Management and Productivity
Chairman: Kenneth S. Cucchiaro, New Jersey Department of Transportation
Chester J. Andres, Steven L. Butts, John F. Direnzo, Grant S. Fairbanks, George E. Gray, William C. Grenke, Ken Hogue, Douglas L. Jonas, Billy M. McCall, William R. Moyer, Marvin E. Mundel, Subhash R. Mundle, David A. Schrier, Jay A. Smith, Jr., Elden G. Spier, Anthony R. Tomazinis, William B. Werther, Jr., Sharon S. Wright

Committee on Strategic Management
Chairman: Kathleen E. Stein-Hudson, Newton, Massachusetts
Terry L. Adams, John W. Barker, John W. Bates, Linda Bohlinger, Ian C. Campbell, Lawrence F. Cunningham, Sidney Davis, James W. Guenther, Irving Hand, Mark P. Howard, George T. Lathrop, Kenneth J. Leonard, Robert G. Maher, Ronald D. McCready, Edwin G. McKeever, James P. Redeker, Noreen Roberts, Aad Ruhl, Peter L. Shaw, Janet Bell Stromberg, Darwin G. Stuart, Bruce D. McDowell

Freight Transportation Section
Chairman: Richard A. Staley, Richard A. Staley Consulting

Committee on Inland Water Transportation
Chairman: Michael S. Bronzini, Pennsylvania State University
Secretary: I. Bernard Jacobson, Temple, Barker, & Sloane, Inc.
George Antle, Kenneth C. Bush, Joedy W. Cambridge, Gary N. Dicer, Gerrit Gort, Lowell D. Hill, Anatoly Hochstein, Kevin H. Horn, William F. Huneke, Marvin L. Jacobs, William F. McFarland, Howard E. Olson, R. Barry Palmer, Craig E. Philip, Larry J. Prather, Miles H. Sonstegaard, Robert N. Stearns, Martin Toyen, Edward F. Tralka

Transportation Systems Planning Section
Chairman: John W. Fuller, University of Iowa

Committee on Citizen Participation in Transportation Planning
Chairman: John N. Balog, Ketron, Inc.
Richard F. Beaubien, Joel P. Ettinger, John W. Fuller, Julie Hetrick Hoover, Leroy E. Johnson, Peter M. Lima, Daniel D. McGeehan, Phyllis Myers, Harold E. Peaks, C. E. Richardson, Marilyn Skolnick, Janet Bell Stromberg, Robin T. Underwood, Kathleen E. Stein-Hudson

Statewide Modal Planning Section
Chairman: Kevin E. Heanue, Federal Highway Administration

Committee on Ports and Waterways
Chairman: Norman B. Wolf, Illinois Department of Transportation
Gordon R. Angell, Pat Broocks-Ross, Carla L. Cefaratti, Wiliam Ronald Coles, Harry N. Cook, Robert Cuellar, Arlene L. Dietz, John W. Fuller, Donald R. Gibson, Herbert R. Haar, Jr., T. Q. Hutchinson, Don H. Jones, Keith L. Kirkeide, John H. Leeper, Paul R. Lowry, Henry S. Marcus, Donald C. McCrory, Warren H. McInteer, William Owusu-Asante, Paul M. Schonfeld, Peter L. Shaw, Harvey E. Sims, Stanley L. Vale

Committee on Transportation Programming, Planning, and Systems Evaluation
Chairman: Thomas F. Humphrey, Massachusetts Institute of Technology
Secretary: Kumares C. Sinha, Purdue University
Gerhard W. Anderson, James H. Banks, William G. Barker, Edward A. Beimborn, Linda Bohlinger, Priscilla S. Cornelio, Joel P. Ettinger, William H. Goldstein, Harvey Haack, Kevin E. Heanue, Gene Lois Hicks, W. M. Hilliard, Monty C. Murphy, Lance A. Neumann, C. William Ockert, Norman G. Paulhus, Jr., Alan D. Pearman, Robert E. Royer, Robert E. Stammer, Jr., William G. Stringfellow, Darwin G. Stuart, D. L. Wieman, Phillip L. Wilson

Kenneth E. Cook, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1987.

NOTICE: The Transportation Research Board does not endorse products or manufacturers. Trade and manufacturers' names appear in this Record because they are considered essential to its object.

Transportation Research Record 1156

Contents

Foreword

The papers in this Record deal with a broad spectrum of issues related to transportation systems planning and management. Also included are several papers dealing with the strategic planning process.

In the first paper, Microcomputer Linear Programming Model for Optimizing State and Federal Funds Directed to Highway Improvements, Theberge develops two network-level models and compares them with existing highway program needs-estimating techniques. Also concerned with highway investment analysis techniques are the papers by Simon, Mackie, May, and Pearman on Priority Assessment Techniques for British Local Authority Highway Schemes and by Berg and Choi on Revision of the Highway Investment Analysis Package Methodology for Estimating Road-User Costs. Berg and Choi propose a revision in the HIAP procedure to more accurately reflect flow conditions in peak-hour traffic. Operating cost data and cost-estimating procedures are modified. Simon et al. compare and evaluate priority assessment techniques currently being used by local transportation authorities in the United Kingdom.

The Highway Performance Monitoring System was designed as a policy-planning tool to predict the operational and conditional effects that highway programs will have in the future. The paper by McPherson and Poole, Use of the Highway Performance Monitoring System To Determine Needs and Travel Cost on North Carolina Highways, provides a general assessment of the North Carolina highway system using the HPMS methodology.

In Sufficiency Ratings for Secondary Roads: An Aid for Allocation of Funds, Mercier and Stoner describe a model that can be used to make sufficiency ratings of secondary roads. The calibrations and scales used with the rating criteria are described in some detail, with emphasis on both the linear and the nonlinear features of the scales.

Solving the Suburban Mobility Problem: Two Case Studies in the Application of Collaborative Problem-Solving Techniques by Bye, Cooper, and Lightbody describes a consensus-building process that can assist stakeholders in reaching agreement and making commitments to implement transportation projects.

Markow, Acharya, McNeil, and Kao assert that life-cycle analysis of waterway facilities requires a new demand-responsive approach to facility performance and to factors that influence costs throughout a facility's service life. Their paper, Management System for Repair, Evaluation, Maintenance, and Rehabilitation of Inland Water Transportation Facilities, sets forth models of facility performance for lock gates, walls, and mechanical equipment. Crew, Hochstein, and Horn also consider water transportation issues. In their paper, Prospects for Container-on-Barge Service on the Mississippi River, they conclude that because of higher transit time for barge shipments compared with rail service, container-on-barge service will not be able to compete for time-sensitive cargoes. Low-value bulk cargoes in containers and return of empty containers may be viable on barges.

Davis, Smith, and Hewa, in Privatization Is More than Contracting Out, distinguish privatization from contracting out, discuss when privatization is necessary, and suggest a four-step approach to implementing privatization.

Three papers dealing with strategic management and planning are included in this Record. Successfully Establishing a Strategic Planning Process by Howard and Initiating the Strategic Planning Process at NJ Transit by Bishop-Edkins and Nethercut describe how public transportation agencies have adapted and applied the strategic planning process that was taken from the private sector. Both papers deal with transit systems serving the New York City metropolitan area. They consider the changing characteristics of and the demand for transit services and the options for services within given financial constraints.

Meyer, in Strategic Management in a Crisis-Oriented Environment, provides a general overview of the process and discusses problems related to its application in a rapidly changing policy environment within the Massachusetts Department of Public Works.

In the final paper, Heuristic Decision Framework for Upgrading Highway Weight Limits, Stephanedes, Ziotas, and Arora develop a method for obtaining a regional road development program that optimizes the net benefits of the projects in a program while meeting specific budget constraints.

Microcomputer Linear Programming Model for Optimizing State and Federal Funds Directed to Highway Improvements

PAUL E. THEBERGE

Funding problems have been one of the major catalysts for implementing and improving the pavement management process. With a drop in anticipated federal funding levels, many states have been and will be required to raise additional state funds. At the same time, states need to recognize the increasing demands for preventive maintenance. Consequently, states need to ensure optimum use of all funds, as well as justify the levels of their requests. To investigate the most effective use of Maine's funding sources, the problem was modeled using linear programming (LP). A simple spreadsheet-based microcomputer LP code that incorporated available pavement management data was employed. The "benefit" measure used to evaluate the variety of available strategies was based on the performance concept of the AASHO Road Test. Actual data on the performance of flexible pavements in Maine were used to construct the models. Two network-level models were developed and applied to the state's Federal-Aid Primary system. One version established "target" miles of the various candidate strategies, and corresponding levels of service, in order to meet a variety of fiscal constraints. The second approach established optimum state and federal budget levels, and corresponding miles of improvement, to meet a variety of performance and resource criteria. The solutions were compared with the needs established by existing methods. They compared favorably and also provided a quick and simple means of evaluating various levels of state and federal budgets. The models also make it possible to support budget requests with objective data.

The pavement management process, introduced in Maine in 1981, initially concentrated on identifying needs to meet a series of top management constraints. More recently it has allowed the department to set priorities among systems, but it has not been used specifically to address the issue of cost-effectiveness of state versus federal programs.

In the Maine Department of Transportation, as in most if not all state highway agencies, recent funding levels have seldom been adequate to meet desired goals. Funding normally comes from both federal and state sources, but the question of how to make optimum use of this combination is seldom addressed.

Federal highway funds are apportioned to states in accordance with a variety of formulas. To obtain these funds, states must match the federal share with state funds in ratios that are dependent on how and where the funds are applied. If states fail to provide their share, the federal funds lapse. On the other hand, state funds used to match the federal share could be used for state-funded activities including maintenance. What are the best uses for these state dollars? State matching funds generate federal funds on a more than one-to-one basis; therefore, it behooves the states to assure optimum use of the resulting funding package. Further complicating the issue is that projected allocations, resulting from the 1987 Surface Transportation Act, will be less than the amount from previous programs. The issue is complex and requires nontraditional forms of evaluation. In any agency, it becomes imperative that an optimization approach be implemented in order to maximize the return on taxpayer dollars.

STUDY OBJECTIVES

The major goal of this study was to develop methodology to examine the optimum use of both state and federal funds in developing a highway improvement program.

The problem was approached by using a "classical" linear programming (LP) approach to model the given objective. LP is a convenient way to model the allocation of resources because it seeks the best solution for a number of competing economic activities.

Introduction during the past few years of microcomputer-applied LP codes has now made it practical and relatively simple to adapt this approach to transportation budget issues. In any operations research study of a specific problem, six phases are generally recognized:

1. Formulate the problem by means of a problem statement,
2. Construct a mathematical model of the system,
3. Derive the solution,
4. Test the solution derived from the model,
5. Establish controls over the solution, and
6. Implement the solution.

This outline served as a guide in addressing the problem presented. Phases 5 and 6, although beyond the scope of this paper, would be required should the findings of this research lead to implementation.

PROBLEM STATEMENT

The Maine Department of Transportation has the choice of applying a variety of repair strategies on many highway sections; the problem is to determine how many miles of each strategy to perform on competing highway sections in order to

Maine Department of Transportation, Box 566, Rockland, Me. 04841.

maximize the level of service provided, subject to the constraints of available federal and state funds, projected deterioration, and other applicable management policy.

STEPS TO A MATHEMATICAL MODEL

The construction of a mathematical model can be initiated by answering three questions:

1. What does the model seek to determine? In other words, what are the unknown variables of the problem?
2. What limitations (constraints) must be imposed on the unknowns to satisfy the model?
3. What is the goal (objective) that needs to be achieved in order to determine the best (optimum) solution for all of the feasible values of the unknowns?

These inquiries were instrumental in conceptualizing the model, formulating the approach, and defining the objectives.

APPROACHES CONSIDERED

The problem that was modeled addressed network-level as opposed to project-level activities. The approaches presented conform to definitions of the department's pavement management system (PMS). The outputs of the models (or unknowns) were designed to address the same goals that the existing PMS addresses at the network or statewide level.

In addition to identifying optimum funding levels or optimum use of limited funding, there were several other associated issues to be investigated. For example, if federal funds were not available, where would the cuts have to be made? If additional state funds could be identified, would they provide more benefit if assigned to state-funded projects or applied to match federal funds? Probably the ultimate question was "how efficient was the latest highway program?"

Two approaches were considered for this study. The first approach was to minimize total expenditures to meet the various resource and service-level constraints established by top management. In reality, this approach might not even lead to a feasible solution depending on the level of management constraints introduced. If that were the case, it would then provide an excellent mechanism for justifying the exact funding levels required to meet management guidelines as well as to establish the amount and source of additional funds required to make incremental improvements in levels of service.

The second approach considered was to maximize the benefits provided under given state and federal budget levels. With this approach, it would be possible to establish the benefits provided by various budget combinations and also to evaluate incremental increases in state, then federal, funding to determine which provided the best return on the dollar. It would also offer an opportunity to evaluate the recently approved program.

The decision variables, or unknowns, under both approaches would be the number of miles of the various strategies to be performed on the different categories of highways. It would then be possible to compare the number of miles with values developed by traditional methods to check for major inconsistencies.

Evaluating Options

Many constraints were common to both approaches. Under the minimizing cost approach, the objective function would be to minimize the total costs of all of the strategies considered. Sensitivity analysis would consist of examining the effects of unit costs on the solution. A second analysis could examine the effect of various service levels and the minimum expenditure needed to attain them. Under the maximization of benefits approach, the budget levels would enter as constraints and the objective function would be the resulting total of projected benefits at the suggested budget levels. Early on, "benefits" were recognized as potentially the weakest component of the models being considered. When constructing a model, it is important to examine the behavior of a solution in response to changes in the parameters of the system. This is especially important when they may be difficult to quantify accurately as is the case for benefits. In this case it is important to perform sensitivity analysis to study the behavior of the solution in the neighborhood of the estimated values. By making "benefits" the objective function, it would be possible to readily perform the required sensitivity analysis of the values employed and thus provide a way of judging the dependability of the models.

Sensitivity analyses could also be performed in the first approach. However, they would not be as convenient and easy to interpret.

Because there were significant advantages to each approach, a decision was made to examine both. In addition, because most of the components were similar, one model form could be easily converted to the other.

Data Requirements

As an aid to understanding the composition and function of the models, the data components are briefly discussed. Actual data were used so the results could be compared directly with those contained in the department's 1988–1989 Highway Needs Report (*1*) developed by the Pavement Management staff. This step is important and conforms to the fourth phase, noted earlier, that calls for checking the models' validity. A common method is to compare the results with some past data for the system being modeled. Because the models were intended to examine actual data, the comparison should reveal favorable results.

Highway Classification Groups

To classify the many miles of highway within each highway system, the existing PMS process aggregates highway sections to reflect various levels of traffic (high and low), standards (adequate and inadequate), and present pavement condition (good to poor). High traffic levels range from 3,000+ average daily traffic (ADT) on the Federal-Aid Primary system to 500+ ADT on the nonfederal state system. Standards of adequacy are based on pavement and shoulder width criteria as well as vertical and horizontal geometrics. The three-variable matrix results in 12 condition states or categories. Table 1 gives this configuration. Because a majority (99 percent) of non-Interstate pavements are bituminous, pavement type is not a variable. The small mileage of rigid pavements is handled on a case-by-case basis.

TABLE 1 HIGHWAY CATEGORIES

Traffic	Pavement Condition[a]	Standards: Adequate	Standards: Inadequate
Low	Good	A1	I1
	Fair	A2	I2
	Poor	A3	I3
High	Good	A4	I4
	Fair	A5	I5
	Poor	A6	I6

[a]Pavement ratings: good = 3.2 to 5.0; fair = 2.4 to 3.2; and poor = <2.4.

Benefit Measures Considered

This item, as noted earlier, introduced the most uncertainty into the models. Four potential measures were considered. Without a doubt, more could have been identified. The choice for this study was made after information obtainable from existing PMS records was considered. This is not to say that the method chosen was the best. As more data are accumulated and analyzed through the department's PMS activities, additional measures could be developed. The areas examined were

1. Reduction in future maintenance costs,
2. Improvement in structural integrity,
3. Reduction in user operating costs, and
4. Improvement in pavement performance.

Each measure was examined in view of how it could be incorporated to reflect benefits of a given strategy as well as provide the necessary measure of demand to which the sum total of all benefits (gain) had to be targeted.

After all of the options had been reviewed and available data were known, a decision was made to employ pavement performance as represented by a measure of a pavement's level of service.

Each of the other options had considerable merit. However, the data necessary for developing the appropriate benefit or demand measures, or both, were either unavailable or questionable. Pavement performance, on the other hand, was selected because there existed historical data on the projected life of the strategies employed in Maine as well the projected loss of service (or deterioration) of the existing highway network. In addition, because of the strong implied relationships between pavement condition and structural integrity and user costs, future correlations would allow those attributes to be examined when relationships became established.

MODEL DEVELOPMENT

After the various budget and system characteristics had been considered, a decision was made that the model or models constructed should be applied to one entire system in the state. It should be noted here that, although this study was of flexible pavements, rigid pavement could be examined using the same approach. The rural Federal-Aid Primary system was chosen because it contained significant mileage (1,826 mi) and had a sizable federal apportionment ($28 million). In addition to the Primary apportionment, substantial surplus Interstate resurfacing, restoration, rehabilitation, and reconstruction (4R) funds were being considered for transfer to this system.

In Maine pavement condition is represented by two measures: distress and roughness. Because historical distress data were more complete and available, distress was chosen to represent level of service. Pavement condition is represented by a rating index of zero to five, with five being perfect. Pavement data collected and tabulated in the spring of 1986 were employed. These are the same data that were used to establish the needs and suggested program levels for the period being evaluated. There are at present eight strategies employed in the programming process. They range from maintenance resurfacing to total reconstruction. Table 2 gives a summary of all of the strategies along with the unit costs and range of life expectancies experienced.

TABLE 2 STRATEGIES, AVERAGE COST, AND RANGE OF SERVICE LIVES

Strategy	Abbreviation	Cost/Mile ($)	Life (years)
Maintenance resurfacing	HHM	12,000	4–6
State light resurfacing	SLR	35,000	6–8
Federal light resurfacing	FALR	77,000	6–8
$1^1/_2$-in. federal overlay	$1^1/_2$ in.	160,000	10–12
Structural overlay	SOL	200,000	12–16
Light federal rehabilitation	RHBL	305,000	14–16
Heavy federal rehabilitation[a]	RHBH	550,000	16–18
Reconstruction	RCN	1,100,000	20

[a]By MeDOT definition, rehabilitation consists of varying levels of improvements within a project, ranging from local base and drainage repairs to reconstruction. Rehabilitation is considered a stopgap form of reconstruction.

The matrix given in Table 3 indicates the combination of highway categories and strategies. Cells in which strategies are not logical or practical have been eliminated and are shown blank. As an example, a federal overlay would not be placed on sections with inadequate standards, nor would a section that met geometric and structural standards be totally reconstructed. This step helped reduce the number of variables significantly from 108 to 48. Also given in this table are the total miles in each category, the average rate of loss or drop in pavement condition for the category, and the total loss of condition for all miles in the category.

Determining Model Coefficients

Under the pavement serviceability concept of the AASHO Road Test (*2*), performance is defined as the accumulated serviceability of a pavement over its life. Figure 1 shows a typical curve indicating past and projected service levels. Graphically, performance is the sum of the area under the performance curve. The projected remaining performance of a section of highway is, therefore, the remaining area until a terminal state is reached. As a pavement section deteriorates, it consumes performance. This represents loss and is represented by the area for a given increment of time. The program period for this study is 2 years. The total projected performance of a highway system at time t is, therefore, the sum of the areas of each individual section from t to t_r. Conversely, the projected

TABLE 3 MATRIX OF STRATEGIES AND HIGHWAY CATEGORIES

	Highway Category[a]											
Strategy	A1	A2	A3	A4	A5	A6	I1	I2	I3	I4	I5	I6
State												
Do nothing	X1	X2		X3	X4		X5	X6		X7	X8	
HMM		X9	X10		X11	X12		X13	X14		X15	X16
SLR		X17	X18		X19	X20		X21	X22		X23	X24
Federal												
FALR		X25	X26		X27	X28						
1½ in.		X29	X30		X31	X32						
SOL		X33	X34		X35	X36						
RHBL								X37	X38		X39	X40
RHBH								X41	X42		X43	X44
RCN								X45	X46		X47	X48
Miles	258	163	69	374	180	72	140	171	95	156	78	71
2-year loss	0.30	0.30	0.60	0.25	0.40	0.55	0.45	0.40	0.55	0.80	0.50	0.75
Total loss	80	50	42	96	72	38	64	69	54	132	40	54

[a]From PMS data base.

loss on a system is the sum of the areas for each section between t and t+2.

Using formulas adopted during the developmental stages of pavement management, the total projected pavement condition, at any time t of each strategy, can be determined using the general form of equation:

$$PCR_t = I - mt^s/20 \quad (1)$$

where

PCR_t = pavement condition at time t,
m = coefficient constant (0.66),
t = time,
s = variable exponent = 1.9/log service life, and
I = rating immediately after improvement.

and the projected performance of an improvement is therefore

$$P = \int_0^L [(I - S_t) - (mt^s/20)]\, dt \quad (2)$$

where

P = performance,
S_t = terminal serviceability value, and
L = service life of treatment.

By using Equation 2, and the projected life expectancies for each strategy combination (X_g to $X48$) of Table 2, performance values were calculated. These values represent the projected area of performance from t_0 *to* t_t as shown in Figure 1. Two values for terminal serviceability S_t were employed. They are 2.2 and 2.8 and correspond to those of Categories A2, A5, I2, and I5 and A3, A6, I3, and I6, respectively. These calculated values of total projected performance or benefits are given in Table 4. Also given is the sum of projected loss in performance for each category:

$$\text{Loss} = 2\{[PCR_t - (L_2/2)] - S_t * M\} \quad (3)$$

where

PCR_t = mean pavement condition rating of category at time t,
L_2 = condition loss for 2 years,
S_t = 2.0, and
m = total miles in category.

Establishing Constraints

In the minimizing cost approach, the aim was to develop the appropriate level of state and federal funding to meet the initial directives of top management. This paralleled the efforts of the PMS process to address the department's needs and also conformed to the results presented in the Highway Needs Report (*1*). Specifically, these directives were that

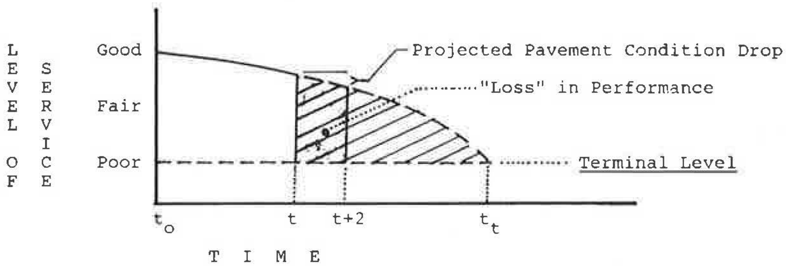
- The total projected deterioration would have to be offset by improvements of equivalent total performance so as to maintain the level of service.
- The average pavement conditions would also be maintained for 2 years.

The PMS process initially identified a level of need referred to as "optimum." The term optimum as used here is intended to mean "preferred" budget level and should not be confused with the definition of optimum as applied to LP solutions. To attain the preferred level, two additional limitations were introduced:

- All deterioration in the poor category had to be offset by primary strategy as opposed to secondary or tertiary options.
- All improvements generated had to guarantee a balanced program over a 20-year analysis period. This was to promote a future steady-state condition by building in the next improvement at the point of terminal serviceability (for example, a strategy with a 12-year projected life is accounted for again in 12 years and not deferred).

In constructing the first model, the four points just mentioned were incorporated as constraints. For future reference, they are labeled in order:

- Performance,
- Pavement condition,



t_o = Initial Time
t = Now
t+2 = Project in Two Years
t_t = Terminal

FIGURE 1 Performance and deterioration concept.

TABLE 4 BENEFITS OF EACH STRATEGY AND CATEGORY COMBINATION

	Highway Category											
	A1	A2	A3	A4	A5	A6	I1	I2	I3	I4	I5	I6
Strategy												
Do nothing												
HMM		4.4	5.8		4.4	3.6		4.4	3.0		2.4	3.0
SLR		5.8	7.6		8.5	9.7		5.8	5.3		4.4	5.3
FALR		10.9	14.5		10.9	11.2						
1½ in.		15.4	21.7		15.4	17.8						
SOL		19.6	26.6		19.6	23.7						
RHBL								17.4	23.7		17.4	23.8
RHBH								19.6	26.8		19.6	26.8
RCN								24.0	33.0		24.0	33.0
Loss of performance[a]	929	196	42	1346	216	38	476	205	54	484	86	54

[a]Total 2-year loss of performance = 4126.

- Deterioration, and
- Project cycle.

Several other constraints also had to be employed to ensure that the model did not generate an unbounded or illogical solution. To start with, the model had to specify the number of eligible miles within the various groups to ensure that the solution would not exceed the limits of available miles. This constraint is referred to as Miles. To model some of the real-world restrictions, it was also necessary to specify a minimum number of miles of reconstruction and a maximum number of miles of resurfacing. The first constraint accounted for committed projects from previous programs for which all preliminary engineering and justifications had been completed. The second constraint accounted for the resource limitations of the department and the paving industry. Because the weather severely restricts the season during which paving operations can occur, finite limits do exist for this part of the highway industry. The values used for this study are based on engineering judgment and historical data. They may be considered arbitrary, but they are real. These limitations are represented by constraints called Min Rcn and Max O'Lay, respectively. Before the model was given its final form, it was recognized that, ideally, a maximum number of miles in the poor categories should be addressed. The number of surplus poor miles in each solution represented miles subject to potential maintenance expenditure. Although this was not, and could not be, included as a constraint, each solution was examined to quantify this potential.

OPTIMIZING BUDGET LEVELS

By employing Tables 2–4, the following model was developed:

$$\text{Minimize budget } Z = Cij\ Xij \quad i = 1, 2, \ldots 9$$
$$j = 1, 2, \ldots 12$$

where Xij is the amount of Strategy i performed in Category j and Cij is cost of Strategy i performed in Category j, subject to the following constraints:

1. Pavement condition: The sum of improvements in condition must equal or exceed the total loss in each category (Table 3).
2. Performance: The sum of improvements in performance must equal or exceed the total loss in performance in each category.
3. Deterioration: The sum of improvements in performance for Categories I3 and I6 must be derived from Strategies 7, 8, or 9 of Table 2.

4. Project cycle: The total number of miles of each strategy for each category times its expected life must equal or exceed the total miles in the category (Table 3).

5. Miles: The total number of miles selected in each category cannot exceed the total number of available miles (Table 3).

6. Max O'Lay: The total number of miles of each resurfacing strategy (Categories 2–6) must not exceed the total available miles divided by the life expectancy in bienniums.

7. Min Rcn: The total miles of Strategies 7, 8, and 9 performed in Categories I2, I3, I5, and I6 must equal or exceed a threshold value (to be entered).

The LP code employed for this study is a spreadsheet-based version called "What's Best!" (3). It was selected because the models could be created in a free format on available spreadsheets using linear formulas. It is based on the well-known LINDO optimization code. The completed model was then applied.

Because this initial version constrained performance within each of the highway categories, it was significantly more restrictive than the traditional PMS approach of total performance only. To examine the effects of relaxing the individual categories (Constraint 2) a series of iterations was run employing percentages (90, 80, and 70 percent) for each category until the total performance gain equaled or approached total performance loss. In some cases it was not possible to reduce total gain to the level of total loss because the performance constraints became nonbinding. After this exercise was completed, another iteration was employed that paralleled the method used in the Needs Study and is referred to as "basic" needs. This consisted of substituting secondary and tertiary strategies for two-thirds of the deterioration in the poor category (Constraints 3 and 7). Even though secondary and tertiary strategies were employed, pavement conditions could be maintained during the 2-year period because more miles were addressed. However, total performance levels were lower.

A summary of the results of both of these exercises is given in Table 5 along with the corresponding values obtained in the Needs Study. The actual miles of each strategy predicted by the various models are given in Table 6 along with those developed in the Needs Study. Two observations are made at this time. First of all, the results obtained, although similar to those in the Needs Study, indicated a significantly higher portion of state-funded work. This is because this model did not differentiate between sources of funds but merely sought the lowest total cost solution. This favored state-funded projects. Second, attention is directed to the small number of miles of 1½-in. overlay predicted by the models. The main explanation for this is that performance loss for Columns A3 and A6 of Table 4 did not turn out to be a binding constraint; therefore, the model chose the lower-cost option (FALR) to minimize cost (X26 and X28 instead of X30 and X32).

MAXIMIZING PROGRAM PERFORMANCE

After the minimizing cost approach was completed, the second approach was employed whereby the objective function was to maximize benefits for given budget levels. This required minor changes in the original model, the first of which was to remove four constraints dealing with meeting performance criteria. This now became the objective function. A second change was made that required adding budget levels as constraints. These changes are represented as

$$\text{Maximize performance } Z = Cij\ Xij \quad i = 1, 2, \ldots 9$$
$$j = 1, 2, \ldots 12$$

where Xij is the amount of Strategy i performed in Category j and Cij is improvement in performance generated per unit of Xij.

Constraints 1 and 3–7 of the original model were retained. The two added constraints were

8. State funds: The total of all miles of Strategies 2 and 3 shall equal some assigned value.

9. Federal funds: The total of all miles of Strategies 4–9 shall equal some assigned value.

The first exercise performed under this approach was to examine the feasibility of attaining a solution employing just federal funds. It was not possible within reasonable limits to obtain a feasible solution. (Feasibility occurred at $85 million.) At this point, state funds were introduced. After repeated runs it became apparent that there existed a minimum sum of state funds with which a reasonable feasible optimum solution could be expected. This value was found to be $3.1 million and resulted in a feasible solution when a federal funding level of $50.0 million was introduced. A reduction to $49 million resulted in infeasibility. At this point it was decided to examine the effects of increased funding. Federal levels up to $60 million and state levels up to $6 million were independently examined. These results are discussed later.

At this point, constraints were relaxed to reflect secondary and tertiary strategies as was done with the original model. Feasibility was attained at $37 million in federal and $3.7 million in state funds. Results of this and the initial run are given in Table 7 along with PMS results. As indicated, these compare favorably. Table 8 gives a summary of the various miles of each strategy suggested by the corresponding options. With the exception of the 1½-in. overlay, there is a striking similarity in the miles of each strategy compared with the corresponding PMS values. One more thing is worth noting about the data in Table 7. Because the traditional PMS analysis did not quantify the levels of performance, it estimated

TABLE 5 COMPARISON OF MINIMIZING COST MODEL OPTIONS

	Budget Level ($ millions)		
	State	Federal/ State	Performance
Optimum PMS	2.5	50.4	N/A
Basic PMS	2.9	37.3	N/A
Model			
Deterioration constraint	7.0	47.5	5365
Match total loss only	4.1	46.5	4733[a]
Relax deterioration	8.6	29.7	4788
Match total loss only	5.8	38.8	4130

[a]Minimum level (performance no longer binding).

TABLE 6 COMPARISON OF MILES OF STRATEGIES, MINIMIZING COST MODEL

			Full Deterioration		Relax Deterioration	
Strategy	Optimum	Basic	Loss by Category	Total Loss	Loss by Category	Total Loss
HMM	168	179	187	226	252	268
SLR	17	17	137	41	160	73
FALR	92	122	117	115	117	115
1½ in.	85	90	42	37	42	37
SOL	13	15	0	0	0	0
RHBL	33	12	63	63	32	32
RHBH	7	4	10	10	3	3
RCN	12	3	17	7	2	3
Total	427	442	573	499	608	530

TABLE 7 COMPARISON OF MAXIMIZING PERFORMANCE MODEL OPTIONS

	Budget Level ($ millions)		
	State	Federal/State	Performance
Optimum PMS	2.5	50.4	N/A
Basic PMS	2.9	37.3	N/A
Model			
Deterioration constraint	3.1	50.0	4897
Relax deterioration	3.7	37.0	4530

the optimum level as status quo when in reality the figure represents a system gain of about 5 percent.

At this point, as with the initial analysis, state and federal funds were incrementally increased independently. The results of that exercise will be discussed later.

To investigate the impact of not specifying a minimum amount of reconstruction, a separate evaluation was performed with Constraint 6 removed. In addition, secondary and tertiary strategies were allowed to provide all of the required performance for Categories I3, I5, and I6. When the model was applied to the $3.7 million state projects level, a feasible solution was derived at a federal level of $35 million. Surprisingly, reconstruction options still entered the solution. This occurred in Categories I3 and I5 so the constraining number of miles of resurfacing would not be exceeded.

All of the solutions from the previous maximization model were examined to determine how many miles in the poor category were not addressed. These were potential miles for some form of major maintenance activity to meet the concerns noted earlier. The data are plotted in Figure 2. Attention is directed to the plot representing the $37 million federal level. As state funding levels were increased from the minimum level, potential maintenance miles actually increased temporarily. This is because the model was maximizing performance and not concerned with the number of surplus miles. Had the maximum number of surplus miles been made a constraint, the solution, for the same number of dollars, would have resulted in a lower value for total performance.

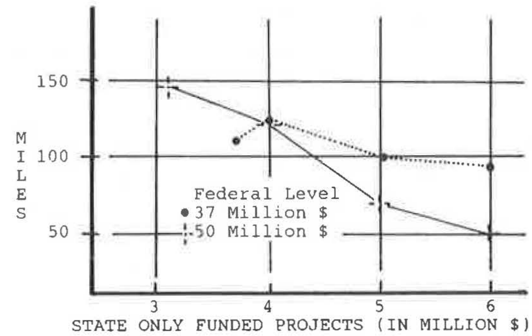


FIGURE 2 Poor miles not addressed at different levels of funding.

DISCUSSION OF RESULTS

Minimizing Cost

The various analyses produced several results worth noting. Portions of every constraint type were influential in each solution. The most prevalent constraint was the one requiring

TABLE 8 COMPARISON OF MILES OF STRATEGIES, MAXIMIZING PERFORMANCE MODEL

Strategy	Optimum	Basic	Full Deterioration	Constraint Relaxed
HMM	168	179	183	211
SLR	17	17	26	33
FALR	92	122	121	121
1½ in.	85	90	19	33
SOL	13	15	29	30
RHBL	33	12	63	40
RHBH	7	4	10	3
RCN	12	3	7	2
Total	427	442	458	473

pavement conditions to be maintained for the 2-year period. Performance criteria were nearly as significant but not necessarily for the same categories of highways. Where performance was not binding, the model chose the lower-cost strategy, unless it exceeded the available supply of miles.

When the performance constraint was relaxed to reflect total performance in lieu of performance in each category, the majority of reductions came at the expense of state-funded improvements. Federally funded improvements could not be cut because they were necessary to maintain average pavement conditions.

To examine the sensitivity of the unit cost coefficient of the 1½-in. overlay on the solution, a series of trials was made at different unit costs. That exercise indicated that unit cost had no effect until it was reduced by more than 30 percent. Even though the strategy provided more performance (profit) it did not enter into the solution because it consumed too much resource (money).

Maximize Performance

When the model was revised, a series of other things became evident. The most significant was that a minimum amount of state-funded projects was absolutely necessary in order to approach meeting the initial constraints with reasonable federal funding levels. Using both the full and relaxed deterioration constraints, it was possible to identify an appropriate level from which incremental increases of both state and federal funds could be evaluated. Figure 3 shows a plot, originating at these minimum levels, of benefits at increasing levels of state funds only, and then federal funds. The plot normalizes each approach by examining total state dollar demands. This was based on a 30 percent state match for federal funds added to the state-funded projects. That exercise was performed to determine the optimum use of additional funds. As can be seen in Figure 3, performance levels were slightly, but not significantly, better when applied to federally funded projects.

A constraint was introduced to specify a minimum number of miles of reconstruction; had it not been, the model would have still specified an amount equal to about 75 percent of the dollar requirement identified in the basic PMS level.

The final analysis consisted of evaluating the actual program. The approved budget for highway-related improvements was $2.6 million of state-funded work and $51.7 million of federal aid projects. This is quite similar to the optimum level suggested. Using the maximization of performance model, and assigning the total miles of each strategy as constraints, resulted in a projected performance total of 4125. This approximates the amount needed to maintain the level of service. It is interesting to note, however, that had the miles of reconstruction been cut back, say to the optimum level, then there could have been an improvement of 5 percent generated with the same funding level.

CONCLUSIONS

The following key points document the success of this investigation:

- Although both models were successful, the maximizing performance model best replicated data from the department's most recent highway needs analysis.
- The maximization of performance model proved to best optimize total state dollars.
- It is possible to document the need for minimum budget levels to meet management constraints with either model.
- It is possible to examine incremental increases in state or federal funding, or both, and to determine the best use of additional funds with the maximizing performance model.
- When the best use of additional state funds was examined, it did not appear to matter (for performance) whether they were applied to state or federal projects. However, program configurations were noticeably different.
- Trying to meet the 100 percent performance requirement for each category with the minimizing cost model appears too restrictive.
- The benefit measure employed, using the performance concept, does not appear to be sensitive to small variations.
- The physical limitations on performing resurfacing strategies has as much effect as any other factors in determining the configuration of a program.
- Major system improvements would require significant increases in resurfacing activities. This could necessitate some changes in policy areas that address the various resurfacing programs.

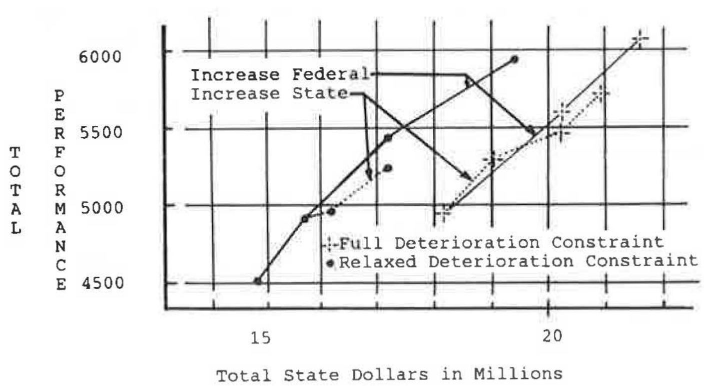


FIGURE 3 Performance for increasing state or federal projects compared with total state dollar demands.

- The low benefit-cost return of the reconstruction strategies suggests that it might be best to perform the analysis in three phases. Phase 1 would optimize all strategies without specifying a minimum number of miles of reconstruction. Phase 2 would optimize the balance of reconstruction strategies to meet management guidelines. Phase 3 would then use the optimum results of Phase 2 as input to re-assess Phase 1.
- The ease with which these models have been able to address a variety of questions suggests that they are far superior to the manual methods now employed.
- The budget as applied in the recent program maintains level of service and is a feasible but not optimum solution for the funds available, if to maximize performance is the objective.

RECOMMENDATIONS

- The models presented here, or some form thereof, should be made part of the department's pavement management process.
- The models should be tested with data from the other major systems.
- The suboptimization approach suggested earlier should be examined as an alternative to arbitrarily relaxing constraints.
- Testing of the models should enable these models, or a variety thereof, to be used in analyzing the needs for the next highway improvement program.
- It is further recommended that, when the next analysis of needs is presented, they include the incremental analyses, as presented here, so that top management can appreciate the alternative approaches that the models present.

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of the Research and Development staff in completing this paper. Thanks are given to Rosanne D'Auteuil and Becky Hunt who provided the typing services for this manuscript. Special thanks go to Brett Porath who provided computer assistance and to Frances Everett for editorial comments.

REFERENCES

1. *Maine's Highway Needs, 1988–1989: An Expanded Approach.* Technical Services Division, Maine Department of Transportation, Augusta, Jan. 1987.
2. W. N. Carey and P. E. Irick. The Pavement Serviceability Performance Concept. *Bulletin 250*, HRB, National Research Council, Washington, D.C., 1960.
3. *What's Best!* General Optimization, Inc., Chicago, Ill., 1985.

The contents of this paper reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Maine Department of Transportation. This paper does not constitute a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.

Priority Assessment Techniques for British Local Authority Highway Schemes

DAVID SIMON, PETER J. MACKIE, ANTHONY D. MAY, AND ALAN D. PEARMAN

Since the mid-1970s, many British local authorities have developed analytical tools to establish a set of priorities among competing highway scheme proposals. These priority assessment techniques (PATs) vary greatly in terms of structure, complexity, data requirements, diversity of schemes to which they are applied, and role within the planning process. Nevertheless they all seek to reduce multivariate information on different projects to a common base, thereby permitting comparison and the setting of priorities in order to optimize the use of scarce capital resources. In this paper, PATs currently used by the local authorities are compared and evaluated, and ways of improving and streamlining their application are suggested. A diverse sample of six PATs is tested on a common set of six highway schemes that have different impacts and costs. The widely different project rankings thus obtained suggest the need for a more homogeneous approach to PAT development and use, and the paper concludes with an outline of a methodology for achieving this.

Local authorities in Great Britain have responsibility for all roads except the 15 030-km motorway and trunk road network, and their highway investment expenditure reached £721 million in 1985–1986. There have, however, been two significant changes in local authority structure in the last 13 years. In 1974 local government underwent a major reorganization, which concentrated highway responsibilities in the hands of a smaller group of English and Welsh counties and Scottish regions. In addition to the existing Greater London Council, six metropolitan county councils were designated in England, covering Greater Manchester, Merseyside, South Yorkshire, Tyne and Wear, West Midlands, and West Yorkshire. Then in 1986 these seven councils were abolished and most of their highway responsibilities devolved to 69 metropolitan boroughs and districts. These changes have certainly had a notable effect on local authority highway investment decision making and capital allocation policy.

The 1974 reorganization brought with it a number of conflicting pressures on these new authorities and the professionals who advised them. Many authorities inherited large highway programs at a time when financial pressures, reaction against highway construction, and concern over blight meant that many highway schemes would not be implemented. Procedures were therefore required for selecting preferred projects from large pools of disparate schemes in wide geographic areas.

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, England.

Three further requirements for any such assessment could be discerned:

1. The need to reflect the wider range of objectives then emerging for transport policy, particularly environmental, planning, and equity issues;
2. The ability to be executed rapidly without undue reliance on complex transport planning models and cost-benefit techniques, which were then increasingly being questioned; and
3. The need for development in consultation with politicians, reflecting the changing relationship between politicians and professionals in the decision-making process.

Several local authorities responded, largely independently, by developing a range of priority assessment techniques (PATs). The purposes for which they were designed included problem assessment, comparison of alternative solutions to problems, coarse sieves to eliminate less urgent or attractive schemes, and establishing detailed priorities among schemes. Although the methods themselves differed, their documentation suggested a degree of uniformity in the criteria on which they were developed (*1–4*):

- Inclusion of the full range of transport policy objectives;
- Assessment of the severity of problems as well as the efficacy of solutions in most cases;
- Adaptability to different kinds of data source;
- Reliance on readily available data;
- Simplicity of execution;
- Adaptability to different value judgments and priorities among objectives;
- Identification of any implicit value judgments; and
- Provision of output as a decision-making guide, not an apparently "right" answer.

These criteria are broadly comparable with those for the Leitch framework, which was being developed for the central government's Department of Transport at the same time (*5–7*) and which also had its roots in multiple criteria assessment methods (*8, 9*). Recent work for the Department of Transport's Urban Roads Appraisal Report (*10*) uses a similar form of analysis but only to examine alternative solutions to a single problem.

In 1986 the Greater London Council and six English metropolitan county councils, many of which had been pioneers in this field, were abolished, but there has been renewed interest among their successor authorities in the shire counties and Scottish regions. Although the Welsh Office has developed a common procedure for use by Welsh counties (*11, 12*), there is still considerable diversity of methods within Britain.

In the United States, many departments of transportation have developed their own schemes for assessing both highway expenditure and public transport support (*13–15*). As is the case in the United Kingdom, and despite some attempts at coordination, a heterogeneous set of systems has emerged. Perhaps because the physical scale and institutional settings are different, the emphasis in U.S. work appears to have been placed on easily implementable sifting systems, intended to eliminate from detailed analysis at an early stage large numbers of schemes with low probabilities of being implemented. Nonetheless, there is a substantial enough core of common interest and shared methodological approach to make cross-fertilization among the approaches employed on each side of the Atlantic potentially valuable.

After earlier exploratory work (*16*) the Institute for Transport Studies has been engaged for the last 18 months on a project sponsored by the Economic and Social Research Council to investigate the range of priority assessment techniques in use; compare their structures, assumptions, and applications; and investigate ways of streamlining their use. In this paper the characteristics of PATs are outlined, a detailed comparative analysis of a sample of methods is discussed, and brief comments are offered on the implications of this exercise for further refinement of such techniques.

PRIORITY ASSESSMENT TECHNIQUES

General Considerations

The PATs developed by local authorities are, with few exceptions, of the points-scoring variety. They are a form of multiattribute decision-making tool designed to facilitate direct comparison of diverse projects. A series of variables relating to traffic, safety, environment, and other relevant issues is defined. These may be measured objectively or subjectively. The evaluating officer gives each problem or solution a score against each variable. These are then summed to produce an aggregate numerical score for each problem or solution. In other words, multidimensional data are reduced to unidimensional form. In addition, each variable or group of variables can be given a weight to reflect policy considerations.

Initial enquiries to 61 English and Welsh counties and Scottish regions drew an encouraging response from 38 (63 percent). Twelve respondents did not use a formal PAT or were only in the early stages of developing one. Two declined to cooperate. Analysis therefore proceeded on the PATs reported by 24 local authorities plus the Welsh Office (*7, 17*). Not all were still in regular use; conversely, a few authorities had developed more than one PAT. Some authorities use the Leitch framework (*5*), the Department of Transport's COBA computer cost-benefit analysis package for major interurban schemes (*18*), or the department's Roads 502 assessment method for small schemes (*19*) instead of, or in addition to, their own PATs. However, these have not always proved appropriate (*7*). Interest in the work of the Institute for Transport Studies was marked; numerous requests for details of the project output were received.

Because of the great diversity of techniques, categorization was attempted using eight possible criteria relating to the internal structure of the techniques and their use within the planning process. Several potentially appealing criteria, such as (a) the cost band of the scheme to which they are applied; (b) whether they are used for inter- or intrabudget ranking; and (c) whether they evaluate problem severity, degree of relief expected, or both, were not used as classificatory tools at this stage because of either insufficient or excessive variation for useful categorization.

Types of Variables

More than 70 percent of the PATs reported have at least some explicitly subjective variables as well as objective ones. "Subjective" means that an officer is required to use judgment in scoring (e.g., decide among categories describing the seriousness of blight).

Number of Variables

With respect to the number of variables in each PAT, several clear categories were distinguishable (Table 1). In practice, this serves as a proxy for the range of information covered because there is a broad correlation between the number of variables and a PAT's comprehensiveness. The methods of the local authorities listed as examples on the right of Table 1 were selected as a representative sample for more detailed comparative analysis of their structure. The availability of documentation and the willingness of officers in the respective local authorities to provide additional assistance as necessary were obviously also prerequisites.

TABLE 1 NUMBER OF VARIABLES IN PATs

Variable Range	No. of PATs	Examples
1–5	4	Gloucestershire
6–10	10	South Yorkshire, West Sussex
11–15	7	West Midlands
16–20	0	
More than 20	4	Devon, Strathclyde

Points System

All of the PATs list a series of relevant variables, often grouped under headings or sections covering traffic, accidents, environment, planning, development, and so forth, against which measured or imputed data are entered. Three-quarters of the techniques use a points-scoring mechanism to compare problems or schemes. Points can be allocated to each variable either by using set conversion rates or thresholds (e.g., one point per 250 vehicles) or by assigning points to particular categories (e.g., no problem = 0, very severe problem = 5). The latter is mostly used for subjective variables. Point ranges can be either open (unbounded) or closed (fixed) range scores (e.g., 0–4 or 1–5). Again, the latter are commonly used with categorical data or subjective variables.

Weighting System

The purpose of a weighting system is to provide a coherent technical or policy-related basis for discriminating among variables. Sixty percent of the PATs attach weights to at least some variables or sections. In some cases, the weights are attached by officers using their technical judgment of relative importance; in others the weights are derived directly or indirectly from council policy statements in documents such as Transport Policies and Programmes (TPPs) and Structure Plans. Both entire sections and individual variables can in theory be weighted, but this may compound the effect if not designed carefully. For simplicity, the weights are most commonly integers, but fractions occur too. In one case (the Welsh Office's SCRAM technique), weights sum to 1 within and between sections in a tree structure (*12, 20*). This approach has great methodological appeal.

QUANTITATIVE COMPARISON OF SIX PATs

Techniques

Even the six PATs selected for detailed comparison differ significantly in the way they measure and calibrate basic traffic and accident data. Other differences in coverage, structure, and use within the planning process are evident:

1. Gloucestershire (5 factors). Other attributes: points scoring, all variables weighted, objective variables only, evaluates problem severity only, applied to all scheme sizes.
2. South Yorkshire (7 factors). Other attributes: points scoring, some variables weighted, objective and subjective variables, evaluates problem severity and solution efficacy, applied to schemes < £250,000.
3. West Sussex (8 factors). Other attributes: points scoring, all variables weighted, objective and subjective variables, evaluates problem severity only, applied to all scheme sizes.
4. West Midlands (12 factors). Other attributes: points scoring, all variables weighted, objective variables only, evaluates problem severity and solution efficacy, applied to all scheme sizes.
5. Strathclyde (39 factors). Other attributes: points scoring, all variables weighted, objective and subjective variables, evaluates solution efficacy only, applied to all scheme sizes. This PAT does not employ fixed or open score ranges in the manner common to other techniques. It awards only scores of +1, 0, and −1 for significant positive, insignificant or zero, and significant negative impacts of schemes, respectively. This clearly precludes direct relative ranking of projects (i.e., a score of 56 is not necessarily superior to one of 55, as would be assumed using the other techniques). In evaluating outcomes, Strathclyde officials take separate account of the unweighted, weighted, and cost-related scores and the number of sections heads and variables under which individual schemes have scored because two or more schemes with the same total points may well have quite different characteristics and scores on different variables.
6. Devon (43 factors). Other attributes: points scoring, all variables weighted, objective and subjective variables, evaluates problem severity and solution efficacy, applied to schemes > £250,000.

Details of the six techniques are given in Simon (*21*). Some indication of the degree of variation within the set of techniques and of the input requirements of the techniques themselves can be gathered from Tables 2 and 3, which list, respectively, the variables used in the smallest (Gloucestershire) and largest (Devon) of the PATs studied in detail.

TABLE 2 VARIABLES USED IN THE GLOUCESTERSHIRE PAT

Variable	Name	Description
I	Accidents	Number of accidents
II	Traffic performance	Route efficiency
III	Environment	Environmental impact
IV	User restraints	User restraint impact
V	Strategic and planning implications	Implications for strategic access and other planning objectives

PATs in the Planning Process

As is evident from the preceding comments, PATs are used in different ways in the respective local authorities' planning processes. In general, simple techniques (which may be unweighted) are used only for preliminary problem identification; more complex (and usually weighted) PATs have the potential for use in successive stages of the planning process, ultimately producing final or nearly final scheme rankings.

It is important to appreciate the applications of the six PATs considered in detail here. This is most clearly expressed in terms of the activities corresponding to successive planning stages (i.e., problem identification; initial sifting of problems or potential schemes, or both; and detailed evaluation and ranking). Initial sifting characteristically attempts to discard proposals that, for various reasons, stand little chance of implementation. In some authorities the evaluation of alternative solutions to given problems is subsumed in this exercise, although it more commonly forms part of the detailed evaluation stage, along with comparison of the optimal solutions to each problem.

Problem-only PATs, however detailed and sophisticated, can by definition only be used in problem identification and ranking. Solution-only PATs are similarly suited only to sifting projects and detailed evaluation. Logically, therefore, PATs intended for use in all planning stages should incorporate both problem and solution components to yield a measure of how well proposed schemes alleviate the problems. Table 4 gives a summary of each authority's use of its PAT. Some overlap between planning stages may occur in practice.

In some cases schemes may be eligible for grants from the Department of Transport through the TPP system or from the European Community's European Regional Development Fund (ERDF). Availability of these funds may affect the mix of schemes finally implemented.

Comparative Exercise

Notwithstanding these differences and the problems they pose for direct comparison, it was resolved that direct quantitative testing of a sample of schemes should be attempted with the

TABLE 3 VARIABLES USED IN DEVON PAT

Variable	Name	Description
I	Traffic	Estimate 1991 August flow Estimate 1991 April flow Existing August congestion Existing April congestion Network impairment by local traffic Network improvement from scheme Proportion of heavy goods vehicles Route relevance to functional route network
II	Accidents	Personal injury accidents (existing road) Pedestrian accidents Fatal accidents Scheme's estimated personal injury accident reduction Scheme's estimated pedestrian accident reduction Scheme's estimated fatal accident reduction Accident-reducing efficacy of cheap safety scheme
III	Highway characteristics	Existing carriageway's structural condition Degree of deficiency from current design standards Standard of bridges, culverts, and so forth Standard of footways, verges Adequacy of existing pedestrian facilities Deficiency of network continuity due to present situation Extent to which scheme will upgrade to acceptable design standards Degree of network functional improvement with regard to continuity
IV	Environment and conflicts	Scheme's relief of Residential area traffic intrusion Pedestrian/vehicle conflict in shopping and industrial areas Sensitive land use disturbance Detrimental environmental effects Noise levels Parking Community severance
V	Commercial and public transport undertakings	Severity of current public transport delays Scheme's reduction of public transport delays Will scheme allow bus priority system Will scheme contribute to heavy lorry route
VI	Development and economy	Will scheme improve access to existing development Is scheme necessary for future development Will scheme facilitate goods vehicle service to shops Will scheme access future housing development Will scheme improve town center access Will scheme facilitate extracounty communications Scheme's housing take Agricultural impact (including farming land lost) Cost Considerations Scheme cost

TABLE 4 PAT USES

	Problem Identification	Problem Sifting	Solution Sifting	Detailed Evaluation	TPP Submission	ERDF Submission
Devon	X	X	X	X	X	
Gloucestershire	X	X				
South Yorkshire	X	X	X	X		
Strathclyde			X			X
West Midlands	X	X	X	X		
West Sussex	X	X				

object of ascertaining the degree to which these PATs produce similar or different rankings of schemes.

To this end, information on a sample of six diverse schemes was obtained from Strathclyde Regional Council. Significant additional data collection and manipulation proved necessary, however, to ensure compatibility with each PAT. Even if essentially the same variables appear, the PATs frequently use different formats or variable definitions. The schemes include a minor rural junction improvement costing £95,000 (Scheme 4); two urban schemes to improve alignments and relieve pedestrian-vehicle conflict caused by heavy through traffic for £1.1 million and £560,000, respectively (Schemes 2 and 3); two rural bridge realignments and reconstructions costing £456,000 and £780,000 (Schemes 1 and 5); and a £5.8 million town center bypass (Scheme 6).

Results

Table 5 gives the scores and rank order of the six schemes according to each PAT. Bearing in mind the points about comparability made earlier, several general conclusions can be drawn.

Scheme Size and Total Points Scores

From the top part of Table 5 it is evident that, irrespective of PAT structure, variable definition, and use for problem or solution evaluation, large schemes tend to score high points. Thus the bypass (Scheme 6) heads all of the rankings whereas the junction improvement (Scheme 4) performed poorly, ranking sixth, fifth, and fourth (twice each) in all cases. There is greater variation among PATs in the ranking of intermediate schemes.

Cost-Related Scores

When cost considerations are taken into account, however, the extreme rankings are reversed in three cases, and some changes also occur in the intermediate ranks (bottom of Table 5). Devon, Gloucestershire, and Strathclyde actually derive cost-

TABLE 5 COMPARISON OF PAT PROJECT RANKINGS

	Scheme					
	1	2	3	4	5	6
Weighted Totals						
Cost (£ millions)	0.456	1.10	0.560	0.095	0.780	5.80
Devon						
Weighted total	307.6	324.3	504.3	321.1	317.4	1,132.9
Rank	6	3	2	4	5	1
Gloucestershire						
Weighted total	202.7	272.2	463.7	202.8	166.8	1,881.5
Rank	5	3	2	4	6	1
South Yorkshire						
Weighted total	9.6	30.4	37.8	7.2	8.6	696.8
Rank	4	3	2	6	5	1
Strathclyde						
Weighted total	5.0	6.0	9.0	1.0	12.0	28.0
Rank	5	4	3	6	2	1
West Midlands						
Weighted total	13.0	41.0	45.0	22.0	65.0	127.0
Rank	6	4	3	5	2	1
West Sussex						
Weighted total	12.0	23.0	45.1	13.0	20.7	134.7
Rank	6	3	2	5	4	1
Weighted Totals/Cost						
Cost (£ millions)	0.456	1.10	0.560	0.095	0.780	5.80
Devon						
Weighted total/cost	248.6	229.2	394.4	348.9	235.9	650.0
Rank	4	6	2	3	5	1
Gloucestershire						
Weighted total/cost	300.2	259.5	619.7	658.0	188.9	781.3
Rank	4	5	3	2	6	1
South Yorkshire						
Weighted total/cost	21.1	27.6	67.5	75.8	11.0	120.1
Rank	5	4	3	2	6	1
Strathclyde						
Weighted total/cost	11.0	6.0	21.0	21.0	15.0	5.0
Rank	4	5	1	1	3	6
West Midlands						
Weighted total/cost	28.5	37.3	80.4	231.6	83.3	21.9
Rank	5	4	3	1	2	6
West Sussex						
Weighted total/cost	26.3	20.9	80.5	136.8	26.5	23.2
Rank	4	6	2	1	3	5

weighted rankings as a standard part of the PAT procedure although they use different cost measures. The other authorities take cost into account during their decision-making process, using pure scheme cost as does Strathclyde, but not within the PAT structure as such. When their scores are divided by cost, an effect similar to that noted previously for the other PATs is observed, even though in the South Yorkshire case the urban bypass (Scheme 6) actually retains top rank. Once again, it is the ranking of intermediate schemes that varies significantly among PATs.

A sensitivity analysis was conducted that showed that, for any single PAT, scheme rankings vary according to the measure of cost used. Most of the PATs tested use pure scheme cost, but one uses the square root of cost and another a discounted log cost formula. Applying, for example, the Gloucestershire PAT to the sample of six schemes yields the results given in Table 6. In the table, cost-weighted scores and ranks are calculated on the basis of

1. Weighted scheme cost/score,
2. Weighted scheme score/log cost, and
3. Weighted scheme score x [log 141/log cost].

Cost measures 1 and 2 are not sensitive to the units in which cost is expressed because both the rankings and the ratios between the actual scores remain constant. With cost measure 3, however, the use of million pounds creates negative scores in four cases. For normal purposes, pure scheme cost is logically most appropriate in that it yields a benefit-to-cost ratio analogous to NPV/C in cost-benefit analysis (*21*).

Analysis by PAT Application

If the PATs are disaggregated according to the purpose for which they are used, the following points are observed:

- Problem severity: The Gloucestershire and West Sussex PATs appear to accord reasonably well overall, agreeing on ranks 1, 2, and 3, despite some differences between them in the relative scores of schemes. In the cost-weighted rankings there are greater differences, although they agree on Rank 4.
- Solution efficacy: Because Strathclyde is the only PAT in this category, direct comparison is not possible.
- Problem severity and solution efficacy: Devon, South Yorkshire, and West Midlands are the three PATs of this type. They all agree on Rank 1. Devon and South Yorkshire also concur on Ranks 2, 3, and 5, and differ only on the other two.

Once again, however, the relative points scored by the respective schemes differ significantly among the techniques. In cost-related terms, Devon and South Yorkshire agree that Scheme 6 (the bypass) remains first, but they differ on all five other ranks; South Yorkshire and West Midlands agree on Ranks 3, 4, and 5.

Overall, then, it appears that there is some correspondence in rankings among PATs designed to evaluate problems, solutions, or both. However, differences in internal structure, variable definition, and weighting account for significant variation. The degree of correspondence among PATs in each of these categories is approximately the same for cost-weighted scores as for scores excluding cost.

Scheme Size and Distortion of Results

Some distortion of results was expected because several of the schemes included in this sample are out of the design cost range of one or more of the PATs. This is true particularly for the bypass, which, at £5.8 million, is many times costlier than the ceiling of £250,000 for the South Yorkshire PAT. Some of the variables included in the South Yorkshire PAT are clearly geared mainly to the smaller end of the cost spectrum (e.g., with respect to footway deficiency and provision). Conversely, some variables to account for strategic issues appropriate to large bypass-type schemes are not included. It is thus interesting that the rankings obtained with this technique did not differ all that much from those of the Devon PAT, with which it is most directly comparable but which is designed for schemes > £250,000. Although the Strathclyde PAT is not directly comparable, because it measures solution efficacy only, its rankings were compatible with the problem and solution PATs at the extremes; it differed only on the intermediate rankings. Given the potential comparability problems referred to earlier, it is difficult to be more precise here.

PAT Appropriateness and Ease of Use

The exercise also clarified several other issues related to the inappropriateness of certain variables and even PATs as a whole, depending on the nature of individual schemes and the importance of using appropriate variables, points, and weights (*21*). These include

- Gaps or double counting with use of inappropriate variables,
- Appropriate variable definition,

TABLE 6 SCHEMES RANKED BY THE GLOUCESTERSHIRE PAT USING THREE DIFFERENT COST MEASURES

		Scheme					
Cost Measure	Unit	1	2	3	4	5	6
1	£m	444.5	247.5	828.0	2,134.7	213.8	324.4
		3	5	2	1	6	4
2	£k	9.49	8.21	19.59	20.81	5.97	24.71
		4	5	3	2	6	1
2	£m	300.2	259.5	619.7	658.0	188.8	781.3
		4	5	3	2	6	1
3	£k	163.8	192.4	362.6	220.4	124.0	1,074.5
		5	4	2	3	6	1

NOTE: k = thousands, m = millions.

TABLE 7 PERCENTAGE OF POINTS BY HEADING FOR THREE PATS

	Scheme					
	1	2	3	4	5	6
Traffic						
Devon	54	48	36	55	60	25
Gloucestershire	6	16	6	5	11	18
West Sussex	36	37	17	11	90	14
Safety						
Devon	34	40	43	30	29	61
Gloucestershire	35	49	53	35	18	41
West Sussex	64	50	64	56	10	70
Environment						
Devon	0	2	11	0	0	6
Gloucestershire	0	1	28	0	0	24
West Sussex	0	12	19	33	0	16
Planning						
Devon	12	10	10	15	10	8
Gloucestershire	59	33	13	59	72	16
West Sussex	0	0	0	0	0	0

NOTE: None of these PATs included a financial implications section.

- Compound or other difficult-to-interpret variables,
- Combining open and closed point ranges,
- Use of ordinal instead of interval or ratio scale points,
- Compounding of individual variable and section weights, and
- Use of different cost measures.

The 43-factor Devon PAT illustrates well that a comprehensive technique need not be complex or clumsy in practice. It was one of the simplest tested because of the clear layout; definition of variables and categories, even when subjective assessment is called for; and the absence of complex formulas. Much the same is true of the Strathclyde PAT, although it was not used in such a detailed manner because the scheme appraisal sheets were obtained in completed form.

Allocation of Points Among PAT Sections and Objectives

The percentages of points allocated to each scheme under the respective PAT sections, which correspond broadly to highway scheme objectives (i.e., traffic, safety, environment, planning including development, and financial implications), were compared. The PATs were then ranked according to the percentage of points in each section to examine differences in scheme performance by section or objective in the various PATs and to establish whether any systematic bias in favor of or against particular sections emerged in the use of any technique.

There are indeed substantial differences among techniques as indicated by the data in Table 7. For example, Devon awards 54 percent of the points for Scheme 2 to the Traffic heading, as opposed to only 6 percent in the Gloucestershire PAT. Scheme 4 scores 0 percent under Environment with the Devon and Gloucestershire PATs, but 33 percent with that of West Sussex. The Planning scores for Scheme 5 range from 0 percent (West Sussex) to fully 72 percent (Gloucestershire). Overall, it is noteworthy that schemes score poorly on Traffic (<19 percent) with the Gloucestershire PAT, but highly (>50 percent) with the West Midlands PAT (not included in the table) whereas West Sussex awards >50 percent of its points on Safety in five of the six schemes.

CONCLUSIONS AND NATURE OF FURTHER WORK

On the basis of information on PATs provided by 25 authorities, six techniques were selected for detailed and quantitative comparison. It has been shown that the six yield significant differences, not only in both pure and cost-weighted project rankings for the diverse sample of schemes but also in the proportional allocation of points among the major sections or objectives of traffic, safety, environment, planning including development, and financial implications. In many respects, this outcome reflects differences in PAT design and use and local authority policies. The implication, however, is that there is a distinct lack of uniformity and standardization among the methods and procedures used by local authorities for priority assessment of highway projects. Some of the techniques certainly leave something to be desired in terms of their technical properties. Furthermore, a number of authorities have no formal PAT; instead they rely on officers' judgment and the political process.

Although it is inevitable that local authorities have different requirements, it does appear that some standardization of overall approach would be advantageous. This does not imply that all authorities should use identical PATs, variables, and weights. Political judgment in this sphere is rightly a local matter. However, this analysis and discussions with representatives of cooperating authorities have led to the conclusion that there is a good case at least for ensuring that the techniques used possess desirable and broadly compatible logical properties.

The final phase of the project is intended to build on the work already completed by proposing improvements to the theory and practical use of PATs. It is intended to develop a general form of PAT that could be used by any local authority, permit a wide degree of flexibility in use, avoid the logical inconsistencies identified in some existing PATs, and incorporate formalized procedures for allocating weights among variables (22). The intention is that this general PAT should be computer based and, in due course, provide computer graphics output to aid decision making.

At the time of writing, initial decisions are being made about the basic design issues for this general PAT. The issues on which initial design decisions have been made follow.

1. The method should be applicable to both problem and solution assessment.

2. The method should be able to assess the full range of highway construction projects but not, at this stage, traffic management or highway maintenance projects, which are viewed as applications for further development.

3. The method should be able to accommodate virtually the full range of scheme costs; only very minor schemes (£25,000) would be excluded. This is a particularly demanding requirement, and the implications for variables are noted in Item 7.

4. The method should include the full range of variables that are of interest to local authorities and relevant to the decision; this raises interesting issues regarding variables such as traffic flow, which is often employed but in practice is only relevant for its contribution to congestion, environmental intrusion, safety, and similar effects that are measured separately.

5. More generally, the method should avoid double counting effects or treatment of both first-run effects such as poor alignment and second-run effects such as accidents, except where there is a clear case for including both to reflect separate problems.

6. Different variables may be needed for certain aspects of problem and solution assessment; rates may be more important for comparison of problems if differences in scale could otherwise bias problem identification; conversely before-and-after differences in absolute values may be more useful for solution assessment.

7. Variables should be arranged in a hierarchical structure, providing a comprehensive set of variables appropriate to large schemes while permitting assessment of more minor schemes against a smaller set of simpler variables, which still attract the same distribution of weights among headings.

8. Points should be scored on an open-ended scale for objective variables but on a closed scale for subjective variables; consistency will, however, be particularly important for the latter, and the facility will be needed to identify schemes for which the upper end of the subjective scale underassesses the size of the effect or problem.

9. The lower end of the open-ended scales may be assessed more coarsely and judgmentally to avoid the need for detailed evaluation of small schemes.

10. Weights should be determined independently by the individual local authority in the light of its policy objectives and used consistently across all schemes; the method should permit both zero weighting for policy issues that the authority considers unimportant and sensitivity testing to make possible assessment of robustness of schemes against changing policies.

ACKNOWLEDGMENTS

The authors are grateful to the Economic and Social Research Council for financing this project and to numerous British local authorities for their support and cooperation, particularly the six whose techniques were compared in this paper.

REFERENCES

1. R. Ball. The TPP: Assessing Priorities. *Journal of the Institution of Municipal Engineers*, Dec. 1976.
2. R. Ball. Priority Ranking of Schemes within the TPP. Presented at Planning and Transport Research and Computation, Summer Annual Meeting, University of Warwick, England, June 1977.
3. A. G. Bovaird, ed. *Evaluation in the TPP Process.* Joint Unit for Research on the Urban Environment, Aston University, Birmingham, England, 1980.
4. M. Roe. *Problem Identification and Priority Ranking in the West Midlands.* Research Note 12. Joint Unit for Research on the Urban Environment, Aston University, Birmingham, England, undated.
5. *Report of the Advisory Committee on Trunk Road Assessment.* Her Majesty's Stationery Office, London, England, 1977.
6. Standing Advisory Committee on Trunk Road Assessment. *Trunk Road Proposals—A Comprehensive Framework for Appraisal.* Her Majesty's Stationery Office, London, England, 1979.
7. D. Simon. *Transport Priority Assessment Techniques Developed by British Local Authorities: A Synopsis and Review.* Working Paper 230. Institute for Transport Studies, University of Leeds, England, 1986.
8. M. Hill. *Planning for Multiple Objectives: An Approach to the Evaluation of Transportation Plans.* Monograph Series 5. Regional Science Research Institute, Philadelphia, Pa., 1973.
9. N. Litchfield et al. *Evaluation in the Planning Process.* Pergamon Press, Oxford, England, 1975.
10. Standing Advisory Committee on Trunk Road Assessment. *Urban Roads Appraisal.* Her Majesty's Stationery Office, London, England, 1986.
11. P. M. Lewis. Decision Making in the Public Sector. Presented at the University of Wales Economics Colloquium, Cardiff, May 8–10, 1983.
12. P. M. Lewis. *SCRAM: Screen Ranking Monitor; Technical Report.* University College of North Wales, Bangor, 1984.
13. *NCHRP Synthesis of Highway Practice 84: Evaluation Criteria and Priority Setting for State Highway Programs.* TRB, National Research Council, Washington, D.C., 1981.
14. M. R. Poole. *Benefit Matrix Model for Transportation Project Evaluation.* Ph.D. dissertation. North Carolina State University, Raleigh, 1982.
15. M. D. Harness and K. C. Sinha. *Priority Setting of Highway Improvement Projects.* Report JHRP-83-9. School of Civil Engineering, Purdue University, West Lafayette, Ind., 1983.
16. W. F. Scotland. *A Comparative Study of Transport Investment Appraisal Systems.* M.S. thesis. Institute for Transport Studies, University of Leeds, England, 1981.
17. D. Simon. *Variables and Data Requirements of Priority Assessment Techniques.* Technical Note 190. Institute for Transport Studies, University of Leeds, England, 1986.
18. *COBA 9 Manual.* Assessments, Policy and Methods Division, Department of Transport, London, England, 1981.
19. *Roads 502.* Assessments, Policy and Methods Division, Department of Transport, London, England, 1975.
20. D. Simon. *Notes on the Welsh Office's Computerised "SCRAM" Assessment Method.* Technical Note 194. Institute for Transport Studies, University of Leeds, England, 1986.
21. D. Simon. *Comparison of Scheme Rankings Using Six Priority Assessment Techniques.* Working Paper 237. Institute for Transport Studies, University of Leeds, England, 1987.
22. T. L. Saaty and K. P. Kearney. *Analytical Planning.* Pergamon Press, Oxford, England, 1985.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.

Revision of the Highway Investment Analysis Package Methodology for Estimating Road-User Costs

W. D. BERG AND JAISUNG CHOI

The Highway Investment Analysis Package (HIAP) released by the Federal Highway Administration in 1976 was revised so that the methodology for estimating operating speeds and costs would be consistent with the results of recent research. The HIAP procedure for accounting for fluctuations in the daily distribution of traffic was first modified to more accurately reflect flow conditions during the busiest hours of the year. The procedure for estimating average travel time under a given congestion level was then revised to be compatible with the 1985 *Highway Capacity Manual* methodology. Finally, vehicle operating cost data were updated, and the cost estimating procedures were modified to be more sensitive to changes in highway alignment. The new FORTRAN code was then validated and tested.

The Wisconsin Department of Transportation (WisDOT) has been performing economic evaluations of alternative highway improvement projects with the Highway Investment Analysis Package (HIAP), a computer model released by the Federal Highway Administration (FHWA) in 1976 for use as a policy planning tool (*1*). The model incorporates procedures for estimating vehicle operating speeds and costs as part of an economic analysis methodology. Since the 1976 release of the model, new capacity analysis procedures and vehicle operating cost data have been developed and published (*2, 3*). In 1979 FHWA published a modified version of HIAP that facilitates the input of highway needs study data and the updating of vehicle operating cost tables (*4*). However, the basic data and procedures for estimating vehicle operating speeds and costs remained intact.

For the HIAP model to remain a useful tool for highway planning purposes in Wisconsin, it was determined that the methodology for estimating operating speeds and costs should be updated to reflect the results of recent research. The objectives of this research were therefore to

1. Revise the HIAP model by incorporating current highway capacity analysis procedures and recalibrating the methodology for estimating vehicle operating speeds and travel time and
2. Evaluate recently published data on vehicle operating costs and revise the HIAP operating cost data base and estimating procedures.

W. D. Berg, Department of Civil and Environmental Engineering, University of Wisconsin, Madison, Wis. 53706. J. Choi, Korea Institute of Construction Technology, Seoul, Korea.

The research essentially followed the HIAP travel time and vehicle operating cost computational sequence; each step was critically evaluated and modified as appropriate. Subsequent tasks involved the programming of the recommended modifications and the validation of the revised software. Replacement pages were prepared for insertion into the HIAP Technical Manual, Computer User's Guide, and Programmer's Guide.

FINDINGS

The HIAP uses a capacity analysis–based procedure for estimating travel time and vehicle operating costs as part of an overall process for assessing the economic desirability of proposed highway improvement projects. Travel times are calculated as a function of the volume-to-capacity (V/C) ratio during six different periods of an average day. The average speeds associated with these six congestion levels are used to estimate vehicle operating costs for an assumed representative highway alignment. In general, the HIAP should be sensitive to changes in the factors that are input parameters to a highway capacity analysis. A limitation of the model in its original format is that it cannot be used to assess the impact of

1. Improvements to specific horizontal or vertical curves,
2. Improvements to intersections, or
3. Improvements that do not influence average speeds.

The HIAP methodology for estimating travel time and vehicle operating costs for a given set of traffic and roadway conditions involves three components: (a) accounting for the daily distribution of traffic, (b) estimating average travel time under a given congestion level expressed in terms of V/C ratio, and (c) calculating vehicle operating costs for each congestion level. The findings of this research will be discussed in the context of these three components.

Daily Distribution of Traffic

The HIAP model distributes average daily traffic (ADT) into six segments that cover the daily range of congestion levels found on a typical highway. The original data were developed by FHWA from 1971 traffic recorder data supplied by several states. Concern had been expressed by some WisDOT officials that the model was not adequately accounting for conditions during the hours of the year that have the highest traffic volumes. A critical review of the HIAP methodology revealed that this concern was well founded.

The original HIAP methodology assumes that the distribution of traffic during a typical day can be stratified into six congestion levels. Each congestion level is defined in terms of the duration of that congestion level in hours and the fraction of the ADT flowing during those hours. Both of these parameters are specified as a function of the ratio of average daily traffic volume to the hourly capacity of the highway section (ADT/C). A sample average daily traffic distribution is shown in Figure 1. The V/C ratio is calculated as

$$V/C = (ADT/C)\ (FRADT)/(DUR) \quad (1)$$

where

ADT/C = ratio of average daily traffic to hourly capacity,
FRADT = fraction of the average daily traffic flowing during the given hours, and
DUR = total hours of an average day during which the flow condition is assumed to exist.

The average hourly volume during any segment is expressed as

$$VOL = (FRADT)\ (ADT)/(DUR) \quad (2)$$

The highest congestion level predicted by the HIAP is that associated with V/C Segment 6. Using as an example a two-lane rural highway with an ADT/C ratio of 3.5, the embedded traffic distribution data in the HIAP model indicate that the most congested hours of the year would have an hourly volume that was only 7.2 percent of the ADT and that these conditions would exist for 6.1 hr each day, or 2,227 hr every year. When similar calculations were performed for the remaining five congestion levels, a plot of these data as shown in Figure 2 revealed that the daily traffic distributions in the HIAP model do not reflect any significant peaking during the highest volume hours of the year. This is demonstrated in Figure 2 by the graph of hourly volume data from automatic traffic recorder stations

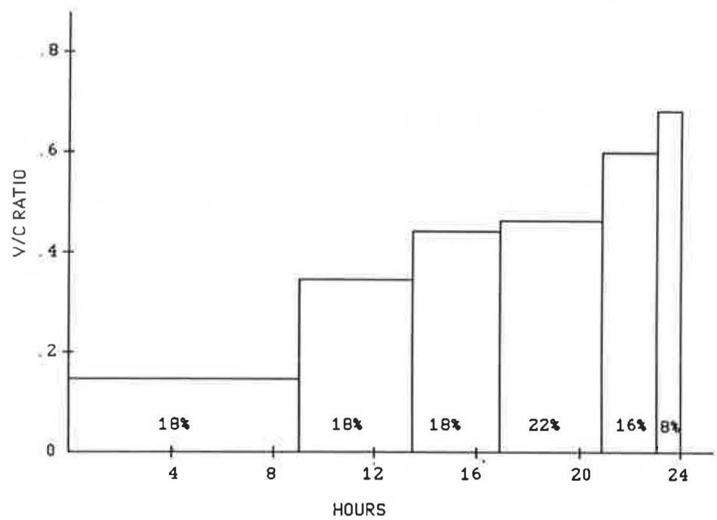


FIGURE 1 Sample distribution of ADT by congestion level and duration.

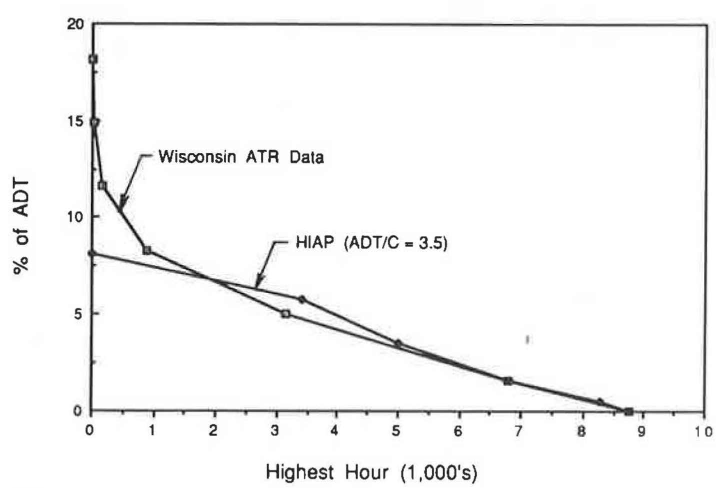


FIGURE 2 Comparison of highest hourly volume distributions: two-lane rural highways.

located on nonrecreational two-lane rural highways in Wisconsin. Similar results were observed for all of the embedded HIAP data.

The implication of these comparisons is that the HIAP model will underestimate the annual road-user costs for a given highway section because of its failure to account for the peaking conditions that occur during the highest volume hours of the year. To overcome this deficiency in the HIAP, the method for estimating the hourly volume and duration (on an average-day basis) of each of the six congestion levels was revised to reflect highest hourly volume distributions found on Wisconsin highways.

The WisDOT Division of Planning and Budget had developed highest hourly volume distributions for six functional highway classifications using automatic traffic recorder data (5). For each distribution, such as that shown in Figure 3, the highest hourly volume expressed as a percentage of ADT was divided into six uniform increments. The midpoint of each increment was defined as the average hourly volume for all hours represented by that increment. Total hours in each increment were determined graphically as shown in Figure 3. The resulting total hours were then factored to an average-day basis to maintain compatibility with other elements of the HIAP model. The complete set of annual traffic volume distribution factors is given in Table 1.

The HIAP source code was then revised to use the Table 1 data as the basis for estimating both the hourly volume and the duration of each of the six congestion levels. The user must specify only the ADT and the factor group that represents the highway under study.

Travel Time

Average travel time on a highway section for each of the six congestion levels that occur throughout a typical day is calculated by the HIAP for each of three vehicle classifications: passenger car, single-unit truck, and multiunit truck. Procedurally, the model calculates the average running speed for automobiles as a function of V/C ratio, the functional classification of the highway, and other design data. This requires that the capacity of each section be either specified by the user or determined internally by the program. Speeds are then modified to reflect added time due to speed changes, stops, and idling. Finally, speeds for other types of vehicles in the traffic stream are estimated on the basis of statistical correlations developed by FHWA in the early 1970s. The resulting speeds are then used to calculate travel time and vehicle operating costs.

Default Calculation of Capacity

Review of the methodology for estimating average running speed revealed that it was developed using data from the 1965 *Highway Capacity Manual* (HCM). The default calculation for the capacity of a rural facility or urban freeway or expressway was

$$C = N \times W \tag{3}$$

where

C = capacity in equivalent passenger cars per hour;
N = 2,000 for a two-lane road,
= 4,000 for a three-lane road, and
= 2,000 × number of lanes for a multilane road; and
W = adjustment factor for lane width and lateral clearance.

For other urban facilities, additional multiplicative factors, given in Table 2, were used.

Because the new 1985 HCM incorporates a number of significant changes to the procedures for calculating the capacity of various types of highways, the HIAP methodology was revised. The new procedures for calculating the capacity of extended sections of freeways, multilane highways, and two-lane highways were included in their entirety with several exceptions. The adjustments for driver population were excluded because their selection must be based on specific local knowledge, and this would be an inappropriate factor in a default calculation. The passenger car equivalencies for buses and recreational vehicles (RVs) were averaged and then used as representative values for the HIAP-defined single-unit truck. The passenger car equivalencies for trucks were used for the HIAP-defined multiunit truck classification.

The HIAP methodology for estimating the capacity of an urban facility was only approximate. Because of the need for a

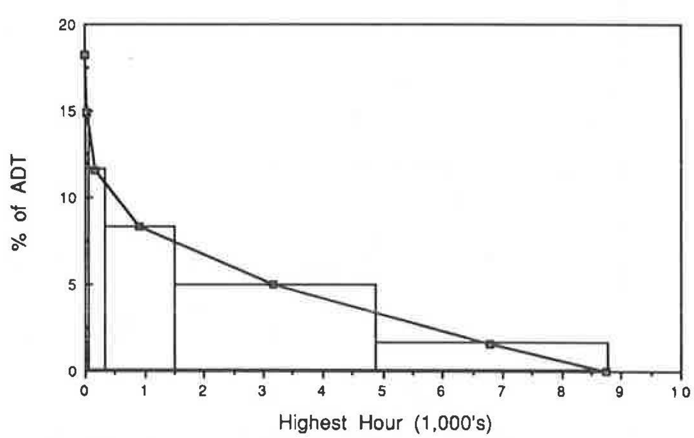


FIGURE 3 Sample annual traffic distribution: Factor Group IV.

TABLE 1 ANNUAL TRAFFIC VOLUME DISTRIBUTION

	Factor Group[a]					
Segment	I	II	III	IV	V	VI
1						
Percentage of ADT	10.3	13.2	14.6	18.2	18.5	18.6
Hours	250	25	25	25	25	25
2						
Percentage of ADT	8.4	10.8	12.0	14.9	15.1	15.2
Hours	750	175	75	25	85	65
3						
Percentage of ADT	6.5	8.4	9.3	11.6	11.8	11.8
Hours	1,550	950	825	225	365	310
4						
Percentage of ADT	4.7	6.0	6.6	8.3	8.4	8.5
Hours	2,150	2,550	2,025	1,225	975	1,000
5						
Percentage of ADT	2.8	3.6	4.0	5.0	5.0	5.1
Hours	1,950	3,150	2,525	3,300	2,975	3,100
6						
Percentage of ADT	0.9	1.2	1.3	1.6	1.7	1.7
Hours	2,110	2,960	3,285	3,960	4,335	4,260

[a]Factor groups are defined as follows: I = urban Interstate, II = urban arterial, III = rural Interstate, IV = rural highway, V = Interstate (recreation), and VI = rural highway (recreation).

TABLE 2 URBAN HIGHWAY ADJUSTMENT FACTORS

	Basic Adjustment	Percentage Green Time Adjustment
Urban central business district	.75	.50
Other urban	.90	.75

procedure that was sensitive to traffic control and geometric design features of stop sign– and signal-controlled intersections typically found within major urban highway projects, a new computational procedure was developed to permit a more realistic estimate of the impact of these facilities on both travel time and vehicle operating costs. The new procedure requires that a signalized intersection capacity analysis be performed by the user before the HIAP model is run. It was determined that the complexity of the new intersection capacity analysis procedures made them inappropriate for incorporation into the HIAP, an already complex model the fundamental purpose of which was to evaluate the economic desirability of proposed highway improvement projects. Directly inputting the performance estimates from the level-of-service analysis into HIAP eliminated the need for coding a default intersection capacity analysis capability. The procedure for accomplishing this will be discussed subsequently.

Average Speed

The HIAP calculates the average running speed of automobiles as a function of V/C ratio. A set of curves is available for a variety of highway facilities. These curves were adapted from the 1965 HCM curves relating operating speed to V/C ratio (*6*). Average running speed as used in the HIAP was defined as the average speed of vehicles on a highway section free of the detrimental effects of internal or external stops or speed change cycles. Because the HCM operating speed was defined as the highest overall safe speed at which a driver can travel on a given highway, it was assumed that the HCM operating speeds were equivalent to the 85th percentile speed and that average operating speeds could be estimated as 90 percent of the 85th percentile speeds. Ratios of running speeds to average operating speeds developed through in-house research by the FHWA were then applied to the new average operating speed curves. The effects of curves and grades, stops, and speed change cycles were estimated separately and then added to the running speed component to obtain the overall average travel speed.

Because the 1985 HCM notes that the curves relating speed to V/C ratio are much flatter than those found in the 1965 HCM, it was decided that the HIAP speed curves should be revised. A factor that also had a significant influence was the use of average speed instead of operating speed in the new HCM. The approach used for revising the speed curves was first to plot the average speed and V/C ratio values defined as the boundary conditions for the various levels of service (see Tables 3-1, 7-1, and 8-1 of the 1985 HCM). These curves were then adjusted at the low V/C ratios to reflect the approximate influence of the 55-mph speed limit. For V/C ratios greater than 1.0, a constant assumed speed for forced flow conditions was established.

Figure 4 shows a comparison of the original HIAP running speed curves for two-lane rural highways with 60-mph design speeds and the revised average speed curves for similar highways. Although the speed values from the old curves would be adjusted to reflect the influence of curves and grades, stops, and speed change cycles, it is evident that there is a significant difference between the two sets of curves. Sensitivity analyses performed using typical data revealed that the revised curves produced results judged to be more realistic and representative of conditions observed on Wisconsin highways. When the revised curves developed for other highway types were compared with the original HIAP curves, patterns similar to those in Figure 4 were also observed.

For urban and suburban facilities where intersection delay and speed limits constitute the principal controls on average

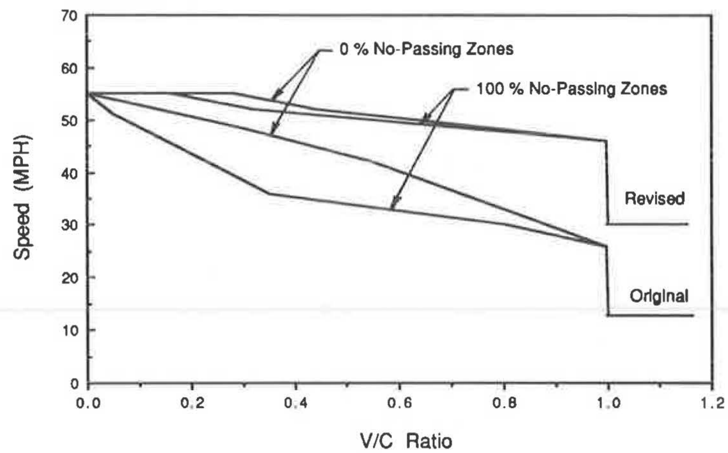


FIGURE 4 Comparison of original and revised speed curves: two-lane rural highway, 60-mph design speed.

speed, there was still a need to be able to account for the effect of the interaction of traffic volume and geometrics on midblock speeds. This was accomplished by using the speed curves appropriate to the cross-sectional characteristics of the urban facility (two lane and multilane divided or undivided) and comparing the resulting value with the user-supplied speed limit. The lower value would be used as the representative midblock average speed.

The original HIAP model also includes two regression equations for estimating truck running speeds as a function of automobile running speeds. Because the revised procedure for estimating speeds begins with a direct estimate of average speed, and because the 1985 HCM makes no distinction among the average speeds of various types of vehicles within a traffic stream, it was assumed that the average speeds from the new curves would apply to all vehicle types. Therefore the regression models for truck running speeds were not retained in the revised procedure.

Curves and Grades

The original version of the HIAP includes a multiplicative adjustment factor to be applied to average running speed as a means of accounting for the effects of average profile and curvature of the type of highway under study. Because the new speed curves were already defined in terms of average speed, the curves and grades adjustment factor was deleted in the revised HIAP methodology.

Speed Change Cycles

The original HIAP model calculates the estimated number of speed change cycles for automobiles, single-unit trucks, and multiunit trucks with a set of regression models developed by the FHWA using a quite limited data base (*7*).

The regression models were formulated by relating automobile speed changes per mile to running speeds. Another model was also developed to relate the magnitude of the speed change to the running speed.

The speed change models for trucks were developed from data that indicated that speed changes varied linearly with truck weight (*8*). The truck data were extrapolated to an assumed 4-kip automobile weight, and ratios between the speed changes for each of the truck classes (single unit and multiunit) versus that for the assumed 4-kip automobile were used as adjustment factors to be applied to the regression models for automobile speed changes.

The resulting set of speed change cycle models is used in the HIAP to calculate an excess travel time that is added to the travel time associated with the previously defined average running speed (as adjusted for the effects of curves and grades). The speed change estimates are also used in the calculation of vehicle operating costs.

Because the new speed curves developed from the 1985 HCM data provide average speed estimates that are assumed to include the effect of speed changes due to traffic stream friction, the speed change models were not used as a component of the revised procedure for estimating the average speed of vehicles under a given set of roadway conditions. However, they were retained for the purpose of estimating vehicle operating costs, with average speed instead of running speed used as the input variable. This was judged to be a reasonable adjustment given that there was substantial dispersion in the original FHWA data.

Stop Cycles

The HIAP model also includes an expression that relates stops per vehicle-mile to automobile speed changes per vehicle-mile. The coefficients in the model vary according to the type of highway facility. Because of the approximate nature of the model and the limited data base used in its development, a revised procedure was designed to more explicitly address conditions that would typically create the need for vehicles to stop.

For purposes of this research, it was assumed that stop cycles occur as a result of the presence of stop sign– or signal-controlled intersections. Neither of these highway features were directly modeled in the original version of the HIAP. This is perhaps the most serious limitation of the original model applied to the evaluation of an urban project. As a result, the HIAP expression for stops per vehicle-mile was only retained for use as one of the required input variables in the equations for estimating truck speed change cycles (discussed in the previous section).

Modeling the number of stops at intersections is a relatively complex procedure. Current methodology is presented in the 1985 HCM. Assuming that the purpose of the HIAP is to estimate user costs and calculate measures of economic desirability, and not to serve as a tool for conducting preliminary design studies, a decision was made not to incorporate the HCM computational procedures directly into the HIAP. Instead, the user would simply input information on both approach delay and stops per vehicle based on capacity analyses or simulation studies conducted before running the HIAP.

In the revised procedure, the approach delay data are used to adjust the travel time derived from the curves relating speed to V/C ratio. The approach delay data are also used to calculate the stopped delay per vehicle, which then serves as the basis for estimating vehicle idling costs. Because the number of stops per vehicle is not normally calculated as part of a capacity analysis, Webster's equation for estimating stops is recommended as an appropriate procedure (*9*).

The influence of intersections on delay and stops depends on the flow conditions and, in the case of a signalized intersection, the traffic signal timing. There are commonly three signal control periods at urban intersections: a.m. and p.m. peak and off peak. During nighttime hours it is also common to have the signals operate in a flashing mode. On rural facilities, there may be only one timing plan, often for an actuated signal controller. To accommodate the HIAP structure of six daily congestion levels, an assumed equivalency between these congestion levels and the possible signal control periods had to be established.

Factor Groups I, III, and V represent various types of Interstate highway, and thus signal control would be nonexistent. For the remaining factor groups, it was assumed that the a.m. and p.m. peak periods each last 2 hr per weekday, that the off peak lasts for 14 hr per weekday and 18 hr on weekend days, and that the signals operate on flash during the remaining 6 hr of each day. Because of the highly directional nature of peak-hour flows, and the need to express the HIAP data on the basis of an average day, it was decided to aggregate the a.m. and p.m. peak data into a single peak period. The values for delay and stops would then be internally calculated by the HIAP as the average of the respective a.m. and p.m. peak-period values. It was further assumed that under flash operation there would be no stops or delay of traffic on the arterial. If cross-street approaches were to be considered, the user would have to code them as separate highway segments.

The resulting distribution of signal control periods is

1. Peak period: 1,043 hr/year,
2. Off-peak period: 5,527 hr/year, and
3. Night period: 2,190 hr/year.

The delay and stop values for these periods are weighted within the HIAP model in accordance with the distribution of hours by V/C segment for the given factor group (Table 1). The weighted values representing each V/C segment are then used to adjust travel time and vehicle operating costs to reflect the presence of stop sign– and signal-controlled intersections. Input coding is restricted to at most one stop sign– and one signal-controlled intersection per highway segment. All vehicle types are assumed to incur the same delay and stops per vehicle.

Idling

The HIAP data for the fraction of time spent idling was discarded so that values consistent with the revised procedure for estimating stops could be used. As discussed previously, data on approach delay per vehicle are available for each V/C segment. Using the methodology of the 1985 HCM, these data are divided by 1.3 to yield an estimate of stopped delay per vehicle. This is then used as the estimated idling time incurred by all traffic on the highway section.

Railroad Crossings

Railroad crossing delays are calculated as the sum of three components:

1. Delays due to trains,
2. Delays due to legally required stops, and
3. Delays due to crossing surface roughness.

A review of the methodology indicated that the estimated relationships for delays due to trains and crossing surface roughness were generally reasonable, although highly approximate. Therefore no changes were made.

However, it was determined that some revisions were justified in the case of delays due to legally required stops. The model for estimating the number of stops included a provision that 24 percent of all crossings equipped with crossbucks were also equipped with stop signs. This assumption was judged to be unrealistic and was therefore omitted. The revised model accounts only for the estimated delay and stops due to trains and to legally required stops by vehicles transporting hazardous cargoes.

Vehicle Operating Costs

Vehicle operating costs are computed in the HIAP over five vehicle types for each of the following components:

1. Running cost (including adjustments for curves and grades),
2. Excess cost of speed change cycles,
3. Excess cost of stop cycles,
4. Idling cost, and
5. Railroad crossing–related costs.

Three basic changes were made to the operating cost methodology. First, recent data developed by Zaniewski et al. (*3*) were used to update the costs to reflect 1980 prices and vehicle characteristics. Second, costs were accumulated over three vehicle classes instead of the original five vehicle types. And, finally, vehicle running costs were made more sensitive to the alignment of the highway section under study.

The Zaniewski data provided the following operating costs:

1. Total cost at constant speed on specific grades and on pavements with specific pavement serviceability indices ($/1,000 mi),

2. Excess cost of speed change cycles ($/1,000 cycles),
3. Excess cost on horizontal curves ($/1,000 mi), and
4. Idling cost ($/1,000 hr).

The data are stratified over three automobile types and five truck types. Because the HIAP methodology for estimating vehicle operating characteristics uses three vehicle classes (automobile, single-unit truck, and multiunit truck), the cost tables were consolidated to approximate the same set of vehicle classes. The weighting factors used in this process are summarized in Table 3 and reflect the average distribution of vehicle types found on Wisconsin highways. The constant speed cost data selected for incorporation in the HIAP were those reported for a pavement serviceability index of 3.0.

TABLE 3 WEIGHTING FACTORS FOR CONSOLIDATING VEHICLE OPERATING COSTS

HIAP Vehicle Class	Zaniewski Vehicle Class	Percentage
Automobile	Automobile	
	Small	26.2
	Medium	21.3
	Large	28.2
	Pickup trucks	24.3
Single-unit truck	Trucks	
	2A single unit	74.7
	3A single unit	24.3
Multiunit truck	Trucks	
	2-S2	10.9
	3-S2	89.1

The other major revision made in the estimation of vehicle operating costs was to incorporate a user-specified distribution of curves and grades for the highway section under study. The HIAP model as presently structured uses a constant assumed distribution of curves and grades for various types of facilities. The operating cost tables in the HIAP were created using these hypothetical alignments. This causes poor sensitivity to assumed modifications in design speed or percentage of highway with passing sight distance greater than or equal to 1,500 ft (both of which reflect the nature of the highway alignment). Although changes in these input parameters will cause a change in highway capacity, and therefore a change in average speed as reflected through the speed–V/C ratio curves, all traffic is still assumed to travel over the same hypothetical alignment for purposes of calculating vehicle operating costs.

In the revised methodology, the user specifies the percentage distribution of a highway segment that has curves and grades within the following four classes:

Class	*Degree of Curve*	*Percent Curve*
1	1–2	Level
2	3–5	± 1–2
3	6–10	± 3–4
4	> 10	> ±4

In this way, an assumed change in highway alignment can be directly accounted for through both the estimated change in average speed and a revised set of cost data to be used in estimating vehicle operating costs. The revised cost data are in the format of Tables 4 and 5. The basic running cost is expressed as

$$BCST = \sum_{i=1}^{p} (WGRD_i)(RGRD_i) \quad (4)$$

where $WGRD_i$ is the fraction of the highway segment with grades within Class i and $RGRD_i$ is running costs at constant average speed on Grade Class i.

TABLE 4 RUNNING COST AT CONTANT SPEED: PASSENGER CARS ($/1,000 VEHICLE MILES)

		Average Speed (mph)													
Class	Grade (%)	5	10	15	20	25	30	35	40	45	50	55	60	65	70
1	level	188	171	151	134	129	126	125	125	128	131	135	141	150	160
2	± 1-2	195	176	153	140	132	127	130	128	130	132	137	142	152	162
3	± 3-4	216	193	168	154	141	132	130	129	132	134	142	148	157	167
4	> ± 4	264	231	201	183	172	161	161	156	158	159	161	163	170	179

TABLE 5 EXCESS RUNNING COST ON CURVES: PASSENGER CARS ($/1,000 VEHICLE-MILES)

		Average Speed (mph)													
Class	Degree of Curve	5	10	15	20	25	30	35	40	45	50	55	60	65	70
1	1-2	3.2	3.9	3.6	3.2	2.8	2.4	2.2	2.0	2.3	4.3	7.4	12.1	18.8	28.1
2	3-5	21.5	20.1	17.5	15.0	10.1	6.2	2.9	0.9	3.3	10.8	24.8	47.8	81.8	136.9
3	6-10	32.3	28.3	19.9	14.2	6.9	5.6	4.0	16.7	45.0	96.1	183.1	219.3	438.1	543.4
4	> 10	32.7	21.1	10.8	4.0	14.5	59.5	112.5	178.6	252.2	381.9	595	--	--	--

The excess cost of vehicle operation on horizontal curves (*ECST*) is

$$ECST = \sum_{i=1}^{p} (WCRV_i)(RCRV_i) \quad (5)$$

where $WCRV_i$ is the fraction of the highway segment with curves within Class *i* and $RCRV_i$ is excess running cost at constant average speed on Curve Class *i*.

Section Detail File

To enhance the usefulness of the HIAP output data, the section detail file was also revised to provide an expanded set of summary statistics for the distribution of the total travel time and operating costs associated with the given highway section. The revisions provide the distribution of these user costs over three classes of vehicle (automobile, single-unit truck, and multiunit truck) and over four operating modes (running, speed change cycles, stop cycles, and railroad crossings). Row and column totals for each 3 × 4 matrix are also provided.

MODEL VALIDATION AND EVALUATION

The new FORTRAN code for the revised HIAP methodology was tested by entering the HIAP computational algorithms into a spreadsheet program for a microcomputer and then comparing the spreadsheet results with those generated by the HIAP program for a set of systematically varied input data. When differences occurred, the FORTRAN code was checked and corrected as appropriate. Although not all table-lookup sequences were fully evaluated, each basic computational sequence was validated. If errors remain, they are probably associated with embedded data values rather than the computational logic.

After the revised HIAP methodology had been validated, the program was applied to five case study highway improvement projects under consideration by WisDOT. The results were then compared with those obtained using the 1979 version of the HIAP. In attempting to generalize the observations from the case study comparisons, it can be noted that the revised version of the HIAP tends to yield lower estimated travel time benefits but larger vehicle operating cost savings. However, this pattern does not hold in all cases. The revised model will produce economic feasibility results that differ from those produced by the old version, but no trend or pattern is apparent.

CONCLUSIONS AND RECOMMENDATIONS

The revisions to the HIAP methodology for estimating travel time and vehicle operating costs provide an improved tool for highway investment planning. The new methodology is now compatible with current state-of-the-art knowledge of traffic flow characteristics and vehicle operating costs. The procedure by which the model accounts for the annual hourly distribution of traffic volume is more consistent with observed data and permits a more reasonable accounting of those conditions that occur during the 100 highest volume hours of the year. These conditions often control the selection of proposed design improvements, especially in the case of highways serving areas with substantial outdoor recreation facilities.

The revised model is also more sensitive to the existing and proposed alignment specified by the user by virtue of the new required input data on the distribution of curves and grades. However, the model still remains insensitive to improvements at isolated locations such as individual curves or grades. Similarly, urban projects or any project that includes a stop sign– or signal-controlled intersection can now be more accurately and reliably evaluated because information on stops and delay is now a required input to the HIAP.

The revised FORTRAN source code has been validated and applied to five case study projects. The results appear to be reasonable. Users of the model are nevertheless encouraged to occasionally check selected calculations to further verify the model or identify elements that require additional testing and evaluation.

In conclusion, the revised version of the HIAP is considered ready for application. It should provide better and more reliable estimates of the economic desirability of alternative improvement projects involving changes in the alignment and cross-sectional design features of extended segments of rural or urban highways. A copy of the revised code and documentation is available from the authors.

ACKNOWLEDGMENTS

This research was funded by the Wisconsin Department of Transportation. Appreciation is extended to Mark Wolfgram, Fred Ross, and Richard Lange of the Wisconsin Department of Transportation for their guidance and assistance.

REFERENCES

1. *Highway Investment Analysis Package.* Report DOT-FH-11-8252. FHWA, U.S. Department of Transportation, March 1976.
2. *Special Report 207: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
3. Texas Research and Development Foundation. *Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Condition Factors.* FHWA, U.S. Department of Transportation, June 1982.
4. *Highway Investment Analysis Package.* Report FHWA-PL-79-014. FHWA, U.S. Department of Transportation, June 1979.
5. B. Aunet and D. Schaul. *Peak and Design Hourly Traffic Volume Factors for Wisconsin State Trunk Highways.* Division of Planning and Budget, Wisconsin Department of Transportation, Madison, May 1981.
6. *Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965.
7. *Highway User Investment Study: Working Paper on Average Running Speed, Speed Change Cycle, and Stop Cycle Relationships.* Statewide Highway Planning Division, FHWA, U.S. Department of Transportation, undated, 25 pp.
8. M. Kent. Fuel and Time Consumption Rates for Trucks in Freight Service. *Bulletin 276,* HRB, National Research Council, Washington, D.C., 1960.
9. *A Manual on User Benefit Analysis of Highway and Bus-Transit Improvements.* AASHTO, Washington, D.C., 1977.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.

Use of the Highway Performance Monitoring System To Determine Needs and Travel Cost on North Carolina Highways

LARRY W. MCPHERSON AND MARION R. POOLE

Providing current information about the performance of an existing highway system has become an important engineering task. Another perhaps equally important task is to predict the operational and conditional effects that alternative highway policies and programs would have in the future. Analytical procedures within the Highway Performance Monitoring System (HPMS) were designed as a policy-planning tool capable of accomplishing tasks of this nature at the state or national level. This paper provides a description of how the HPMS was applied in a statewide study to (a) establish relationships between investment levels and performance of the existing highway system, (b) estimate highway needs for that system over a 10-year analysis period, and (c) estimate future highway system user costs as a function of investment. There was no attempt to critique either the function or the philosophy of the HPMS. The study results reported in this paper constitute a general assessment of the North Carolina highway system that verifies or identifies, or both, highway statistics, deficiencies, and needs over time. Many of the study findings should provide indispensable information to North Carolina highway administrators and decision makers. On the basis of those findings it was concluded that the HPMS can be used to quantify program needs that, if met, would lead to optimal achievement of a state's highway transportation goals.

In this paper are described several aspects of the North Carolina highway system including its physical condition, operational characteristics, and roadway needs from 1983 through 1993. The focus is on the performance and condition of the existing arterial and collector systems. Local roadway needs were not analyzed because the data base used for the study was limited to Interstate, arterial, and collector highways within the federal functional classification system. Data base limitations also precluded an assessment of bridge needs and requirements for highway construction at new locations. However, these needs have been determined and documented in other studies (*1*).

The Highway Performance Monitoring System (HPMS) is the primary source of information reported here. This data base system supplies timely information about the condition and use of major highway systems and about the capital investments being made to improve them. It can provide personnel involved in highway program development and management with a continuous view of how and where conditions are changing and how investment levels and patterns relate to those changes. The HPMS was developed and implemented jointly by the Federal Highway Administration and state highway agencies. These agencies currently support and maintain this data base system.

Planning and Research Branch, Division of Highways, North Carolina Department of Transportation, P.O. Box 25201, Raleigh, N.C. 27611-5201.

The primary investigative tools used in this study were the analytical models and the inventory data base that are couched within the HPMS environment. Two types of data make up the HPMS inventory. They are generally referred to as (a) universe data and (b) sample section data.

Universe data define the extent of roadway mileage by functional system and jurisdiction. Over and above the universe data are sample section data that are routinely collected on randomly selected sections of the arterial and collector highway systems.

The sample sections are spatially fixed and have homogeneous geometrics. These sample sections were selected in accordance with the *Highway Performance Monitoring System Field Manual* (*2*) and provide the physical and operational data base from which the performance of the highway system can be evaluated.

Several HPMS models were used to accomplish the analyses required for this study. Those models were designed to analyze the sample section data and establish relationships between various levels of capital investment and the resultant performance of the highway system. Study references provide a complete description of these models ranging from an overview of their use for obtaining highway performance information to a detailed discussion of their analytical potential as policy-planning tools (*3–7*).

PRESENT HIGHWAY TRAVEL, CONDITION, AND PERFORMANCE

Definitions

The performance of a highway system is defined as the degree to which the system serves the movement of people and goods safely, efficiently, and economically. A highway performance measure is defined as an indicator of highway service derived from the condition, usage, operation, and physical characteristics at a particular time (i.e., past, present, or future). Important examples of highway performance measures are peak-hour operating speed, volume-to-capacity ratio, pavement condition, roadway cross sections and alignments, system mileage and travel, accidents, and user costs. In this paper are reported some of the more important performance indicators for North Carolina highways, which existed on December 31, 1984, as derived from the HPMS data base. These data are updated annually and are the primary source of information about the physical condition, extent, and usage of the state highway infrastructure (*2*).

Highway Mileage and Travel Estimates

The HPMS estimates of mileage and travel on North Carolina's 1984 highway system are given in Table 1. Data in this table are stratified by the federal functional classification system (*8*) and show that there were 120,083,000 daily vehicle-miles traveled (DVMT) distributed over 92,719 highway miles. It should be noted that the distribution of travel is not directly proportional to mileage. The data in Table 1, for example, indicate that rural highway mileage is nearly 81 percent of the statewide total but carries only 58 percent of the travel. The data also indicate that Interstate highways are less than 1 percent of the highway system yet carry more than 14 percent of the total travel.

The percentages of rural and urban travel are expected to change in the next few years. It is anticipated that urban travel will increase while rural travel decreases. These changes will be due primarily to redefinition of rural areas as urban areas in the 1990 census. The HPMS will be useful in tracking such areal changes in mileage and travel over time.

Highway Performance Relative to Condition, Safety, and Service

Performance of the North Carolina highway system is related to many physical and operational characteristics. Some of these characteristics are relatively fixed (e.g., lane width and alignment) and some can change rapidly (e.g., pavement condition).

HPMS data that can be used to define performance are organized in the three broad categories of condition, safety, and service. Condition data include information on pavement type, pavement condition, and drainage adequacy. Safety data are information on roadway cross-sectional (i.e., lane, shoulder, and median widths) and alignment adequacy. Service data include information on operating speed, volume-to-capacity ratio, and access control. The HPMS analytical models provide information on highway system performance by measuring and reflecting changes in the system's performance indicators.

Condition

Table 2 gives a summary of 1984 pavement condition by functional class. Pavement condition is defined by a present serviceability rating (PSR) code. The range of PSR codes and their meanings are given elsewhere (*2*). It must be pointed out, however, that the derivation of the PSR (as a measure of pavement condition) is not rigorous. It is based on the engineering judgment of state highway personnel and norms acceptable to those engineers.

A pavement with a PSR of 2.5 or lower has deteriorated from a good or fair condition to a point where resurfacing or pavement rehabilitation is needed. The data in Table 2 indicate that the lower functional class systems tend to have a greater percentage of their highway mileage in this PSR category (e.g., 38.3 percent of the urban collector system) than do the higher functional class systems (e.g., 5 percent of the urban Interstate system). The first column of Table 2 shows an all-functional-class total of 6,044 highway miles that needed either resurfacing or pavement rehabilitation in 1984 (*8*).

Safety

The physical features that contribute to overall driving safety on the highway can be specified as either geometric or cross-sectional. Geometric features are the elements used for the roadway's horizontal and vertical alignment. Lane width, shoulder width, divided roadways, or undivided roadways are important cross-sectional features for highway safety.

TABLE 1 NORTH CAROLINA HIGHWAY SYSTEM MILEAGE AND TRAVEL IN 1984

Functional Class	Miles	Percentage of Total Miles	DVMT	Percentage of Total DVMT
Rural				
Interstates	595	0.6	10,993,000	9.2
Other principal arterials	2,014	2.2	12,609,000	10.5
Minor arterials	1,987	11.4	21,869,000	6.4
Major collectors	9,173	9.9	6,750,000	5.7
Local	50,771	54.8	9,900,000	8.2
Subtotal	75,101	81.0	69,846,000	58.2
Urban				
Interstates	201	0.2	6,311,000	5.3
Other freeways and expressways	209	0.2	4,996,000	4.2
Other principal arterials	1,641	1.8	19,373,000	16.0
Minor arterials	2,125	2.3	10,619,000	8.8
Collectors	1,330	1.4	2,289,000	2.0
Local	12,112	13.1	6,649,000	5.5
Subtotal	17,618	19.0	50,237,000	41.8
Total	92,719	100.0	120,083,000	100.0

TABLE 2 NORTH CAROLINA HIGHWAY SYSTEM 1984 PAVEMENT CONDITION MILEAGE GROUPED BY PSR

Functional Classification	PSR 2.5		PSR 2.5–2.9		PSR 3.0–3.4		PSR 3.5–3.9		PSR 4.0		Total	
	mi	%	mi	%	mi	%	mi	%	mi	%	mi	%
Rural												
Interstates	16	2.7	49	8.2	168	28.2	207	34.8	155	26.1	595	100.0
Other principal arterials	309	15.4	307	15.2	135	6.7	244	12.1	1,019	50.6	2,014	100.0
Minor arterials	403	20.3	324	16.3	122	6.1	411	20.7	727	36.6	1,987	100.0
Major collectors	2,108	20.0	1,828	17.3	447	4.2	1,754	16.6	4,424	41.9	10,561	100.0
Minor collectors	2,021	22.0	2,250	24.5	860	9.4	987	10.8	3,055	33.3	9,173	100.0
Urban												
Interstates	10	5.0	25	12.4	39	19.4	57	28.4	70	34.8	201	100.0
Other freeways and expressways	21	10.0	53	25.4	15	7.2	15	7.2	105	50.2	209	100.0
Other principal arterials	343	20.9	170	10.4	164	10.0	193	11.8	771	46.9	1,641	100.0
Minor arterials	311	14.6	237	11.2	272	12.8	274	12.9	1,030	48.5	2,124	100.0
Collectors	502	38.3	109	8.3	136	10.4	238	18.1	327	24.9	1,312	100.0
Total	6,044	20.3	5,352	17.9	2,358	7.9	4,380	14.7	11,683	39.2	29,817	100.0

Divided highways with full access control eliminate most cross traffic conflicts and provide unlimited passing opportunities. Elimination of traffic conflicts explains why divided highways have the lowest fatality rates.

Important features that reduce the potential for accidents on undivided facilities are adequate lane and shoulder widths, proper passing sight distances, and good stopping sight distances.

The HPMS models output both highway mileage and travel by cross-sectional type and functional class (*5*).

Service

Operating speed and levels of congestion tend to be the best indicators of the service component of highway performance. Operating speed is primarily a function of congestion levels in urban areas. Traffic congestion will also limit operating speeds in rural areas but not as much as curves and grades. Levels of congestion depend on highway traffic volumes and capacity.

The service component of highway performance is also quite sensitive to changes in travel growth rate. Indeed, traffic growth can have a significant impact on aggregate highway performance. This impact can become rather acute over the short range because it usually takes a long time to plan, finance, design, and construct highway capacity improvements.

Service component indicators output by the HPMS models include (a) expected traffic growth rate, (b) levels of peak congestion, and (c) speed of trip making during peak periods (*5*).

NORTH CAROLINA HIGHWAY NEEDS ESTIMATE (1984–1993)

Determinants of Highway Needs

The three interrelated variables of present conditions, future travel, and investment levels will determine future highway needs and conditions. These variables are used in HPMS analyses as a basis for establishing investment-performance relationships. By selecting appropriate minimum tolerable conditions (MTCs), types of construction improvement, design standards, travel projections, and funding strategies (*5*), the HPMS user can tailor analyses for the evaluation of specific policies or situations. The HPMS analyses accomplished during this study were tailored to (a) estimate total rural and urban highway needs through 1993, (b) yield relationshipsbetween various levels of capital investment and system performance, and (c) determine the 1993 cost of highway travel in North Carolina relative to three different levels of investment.

Highway Needs Estimate Through 1993

Assessment of needs is the first step in investment-performance analyses. The HPMS defines highway needs in terms of the funding level required to maintain a highway system at or above certain MTCs. Dropping below the chosen MTC-values implies a state of deficiency. The level of funding necessary to correct all deficiencies as they occur is called full needs funding. The needs model determines full needs funding by first identifying deficiencies and then simulating the type and cost of capital improvements required to correct those deficiencies. Such a funding level estimate is objectively based on a cost to maintain the highway system's level of performance defined by MTC-values. It should be noted that HPMS needs assessment is accomplished without regard to revenue availability, user cost distribution, jurisdictional responsibility, or other subjective factors that actually determine highway program direction and investment levels.

Three major types of look-up tables are required by the needs model. These tables contain MTC-values, design standards, and costs for both right-of-way and construction. System default values and standards are national averages. System default values were used for the needs analysis phase of this study. The results of that analysis are shown in Figures 1 and 2, and further data are given in Tables 3 and 4. Costs are in 1981 dollars.

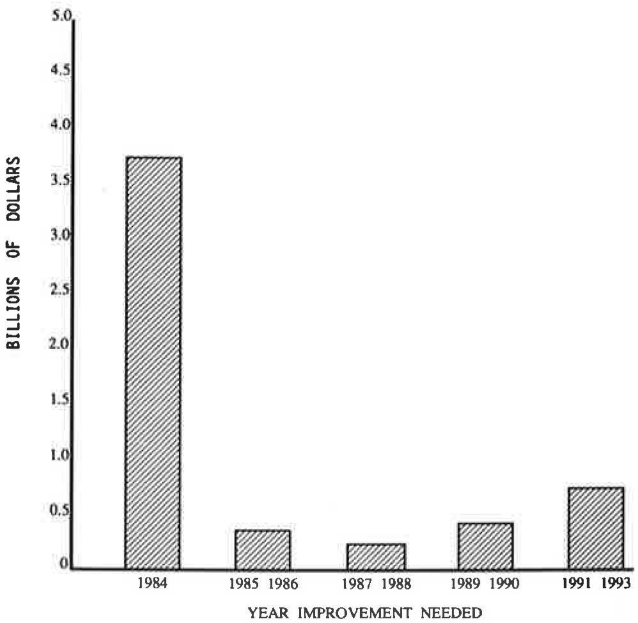


FIGURE 1 Rural highway needs estimate, 1984–1993.

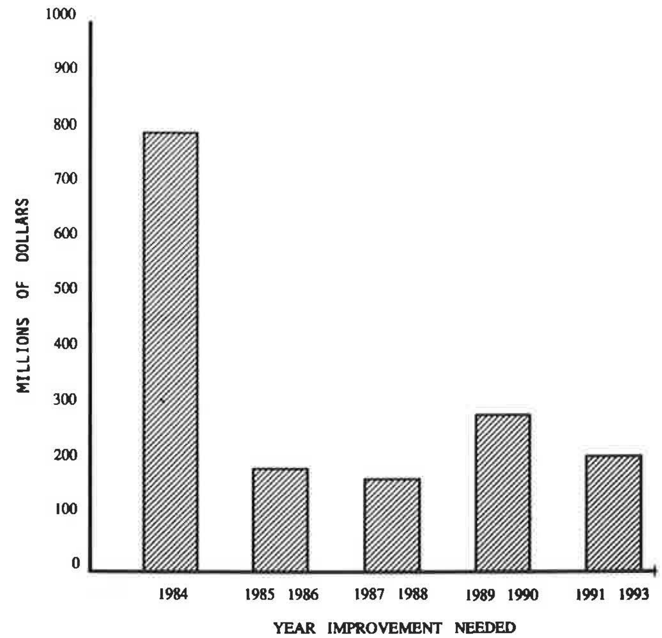


FIGURE 2 Urban highway needs estimate, 1984–1993.

TABLE 3 RURAL COST ($000, 1981)

ALL FUNCTIONAL CLASSES

	1984-1984		1985-1986		1987-1988		1989-1990		1991-1993	
	MILES	COST	MILES	COST	MILES	COST	MILES	COST	MILES	COST
RECONSTRUCT TO FREEWAY	157	351071	29	65821	51	114794	86	193417	55	122744
RECONSTRUCT W/MORE LANES	34	72442	8	19073	0	0	34	80953	8	19068
RECONSTRUCT W/WIDER LANES	1401	1384750	77	78657	0	0	0	0	27	19281
PAVEMENT RECONSTRUCTION	691	599359	0	0	0	0	0	0	15	9821
ISOLATED RECONST (ADD LANES)	2181	571936	137	27212	0	0	0	0	0	0
MAJOR WIDENING (ADD LANES)	107	173814	15	21454	21	26142	56	99889	126	231181
MINOR WIDENING	1137	256274	68	14738	0	0	1	284	266	58530
RESURFACING W/SHLDR IMP	286	65393	108	22383	82	13768	94	16888	168	30504
RESURFACING	910	63614	1293	105830	693	53014	1691	92583	2047	109424
RESURF W/ALIGN & SHLDR IMP	76	60278	0	0	0	0	0	0	0	0
RESURFACING W/ALIGN IMP	66	29008	0	0	0	0	0	0	0	0
TOTAL	7047	3627939	1735	355169	847	207719	1962	484015	2711	600553

TABLE 4 URBAN COST ($000, 1981)

ALL FUNCTIONAL CLASSES

	1984 - 1984		1985 - 1986		1987 - 1988		1989 - 1990		1991 - 1993	
	MILES	COST	MILES	COST	MILES	COST	MILES	COST	MILES	COST
RECONSTRUCT TO FREEWAY	0	0	0	0	0	0	0	0	0	0
RECONSTRUCT W/MORE LANES	24	149641	0	0	0	0	0	0	0	0
RECONSTRUCT W/WIDER LANES	4	6108	0	0	0	0	0	0	0	0
PAVEMENT RECONSTRUCTION	108	160861	12	17593	2	2840	25	84929	49	53429
MAJOR WIDENING (ADD LANES)	21	115713	9	62814	8	50370	14	110589	10	37855
MINOR WIDENING	18	14434	1	4513	0	0	0	0	0	0
RESURFACING W/SHLDR IMP	224	80689	58	16322	20	6020	19	6638	52	22042
RESURFACING	730	270361	272	92635	305	102076	249	88709	366	111325
TRAFFIC ENGINEERING	4	176	1	42	1	29	8	81	1	29
TOTAL	1133	797983	353	193919	336	161335	315	290946	477	224680

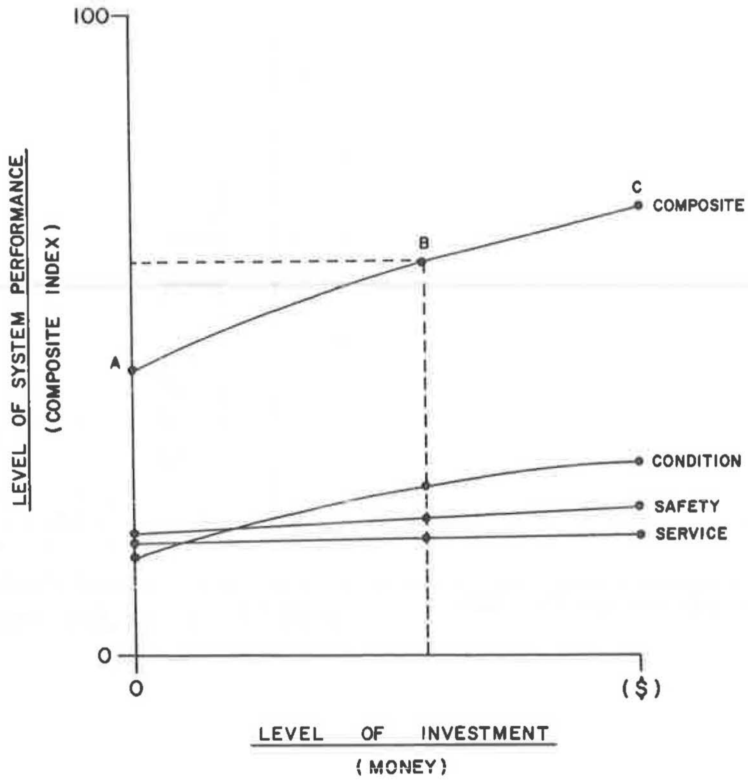


FIGURE 3 Investment-performance relationship (*4*).

NORTH CAROLINA HIGHWAY TRAVEL COST ESTIMATE (1984–1993)

Development of Investment-Performance Relationships

Highway investment-performance relationships output by the HPMS analytical procedure are based on theoretical and empirical research conducted by federal and state governments, AASHTO, and several leading universities (*4*). These relationships, as shown in Figure 3, permit estimates of future performance of a highway system given investment patterns and levels, future travel estimates, and applicable design standards. Such investment-performance curves can be developed for each functional class system from the HPMS investment-level analyses.

The two types of investment-performance analyses that the HPMS can accomplish are known as "investment level" and "funding period" (*5*). Both types were required during this study. Investment-level analysis was used to determine total highway needs and to estimate base year conditions and vehicle performance impacts. Funding-period analysis was used to forecast target year conditions and vehicle performance impacts.

It should be noted that either base or target year conditions can be analyzed by the impact model but that target year conditions and impacts can be analyzed only during a funding-period analysis.

The HPMS investment-level model simulates seven funding levels ranging from full needs investment to no investment at all. The full needs investment level (or 100 percent funding level) simulates highway system effects for all improvements selected by the needs model. The next six funding levels have the respective percentages of 80, 70, 60, 40, 10, and 0 percent of full needs investment. The lower funding levels simulate only a portion of total required improvements, depending on the relative amounts of funds available. The zero funding level simulates the effect of making no capital improvements during the analysis period, which, for this study, was 10 years. The zero funding level is represented by Point A on the composite curve in Figure 3 and the 100 percent funding level is represented by Point C.

After conducting the investment-performance analyses, the HPMS model outputs seven points for each of the safety, service, condition, and composite curves. The shapes of these curves are discretely and uniquely determined by those point sets.

The graphs are developed by plotting the composite index values versus the dollars funded for each level. These graphs can be used to provide answers to many highway programming and budgeting questions. For example, management may desire an estimate of the budget level needed to maintain current or base year conditions on the highway system. The dotted lines in Figure 3 illustrate how that estimate can be obtained. Similar intercepts for other desired levels of performance would yield different investment estimates depending on the slopes of the curves.

Investment-performance graphs were developed for each functional class within the North Carolina highway system.

Changing Highway Conditions and Resultant User Cost Impacts

Changes in pavement conditions, traffic congestion levels, operating speeds, and roadway geometry affect the costs of using the North Carolina highway system. Costs for using that system can be estimated from data output by the HPMS simulation procedures (*5, 9*). Those procedures simulate the way that highway conditions affect vehicle performance. Simulation results are expressed as vehicle performance indicators that include speed, fuel consumption, operating costs, emissions, and accident rates. With the exception of accident rates, all of the performance measures are summarized by vehicle type.

Vehicle performance indicators provide a flexible means of comparing cost of travel estimates for different highway program, policy, or investment strategy scenarios. For example, vehicle performance indicators can be converted to user cost units and (a) compared for base (or existing) and target (or forecast) years to obtain the effects of a single program over time or (b) compared at the target year for several alternative programs to obtain relative cost of travel differences.

For purposes of comparison, the HPMS analytical procedures produce highway travel cost components for both base and target years. Cost components reported by these procedures are (a) average travel speed, (b) accidents, and (c) operating costs. The respective measurement units for these components are miles per hour, accidents per 100 million vehicle miles traveled (VMT), and dollars per 1,000 VMT. These units can be equated to a single unit of money and subsequently combined to yield an economic basis for comparing alternative highway programs.

The task of assigning monetary values to the variables of travel time or accidental death and injury is usually quite difficult and subjective in nature. This task must be accomplished, however, before any analysis designed to yield total user cost differences among highway program alternatives can be conducted. Accident and travel time unit cost values used in this study are given in the following tables (*9*):

Type of Accident	*Unit Cost ($/accident)*
Damaged vehicle	471.00
Nonfatal injury	3,854.00
Fatality	268,727.00

Type of Vehicle	*Unit Cost ($/hr/vehicle)*
Light automobile	5.56
Heavy automobile	5.56
Pickup and van	6.78
Single unit (2 axle)	7.51
Single unit (3+ axle)	10.00
Multiunit (4 axle)	11.00
Multiunit (5+ axle)	11.00

Cost of Travel as a Function of Investment

The final objective of this study was to estimate the 1993 cost of travel on the North Carolina highway system for different levels of funding while all other variables were held constant. This objective was accomplished in two steps. First, vehicle performance measures were determined for each funding scenario by simulation. The second step involved calculating the total cost of travel based on the vehicle performance measures determined in the first step. The methodology used to calculate those costs is outlined in the appendix of the 1985 biennial report to Congress by the U.S. Department of Transportation (*9*). An electronic worksheet will easily accomplish the calculations required by that methodology.

Year 1993 cost of travel calculations were conducted for three funding levels. The zero level provided no money for correcting highway deficiencies during the analysis period, 1984–1993. The maintain-conditions level provided only enough money to maintain a 1984 level of service through 1993. The 100 percent level of funding provided money to correct all deficiencies found on the highway system during the time of analysis. A summary of the calculations for the combined costs and unit costs by area type and functional classification for each funding level at the 1993 target year are given in Tables 5–7.

SUMMARY

Throughout the nation there has been a growing recognition of the necessity of periodically assessing the extent, physical condition, efficiency, economy, and safety of the highway system. In addition to such a general assessment, it is sometimes desired to evaluate the economic impacts of various highway programs and policies. A general assessment of the North Carolina highway system during the 1984–1993 period was the main objective of this study. An important secondary objective was to estimate user costs on that system as a function of investment level. These two objectives were accomplished by using the HPMS analytical procedures. Study results and findings should be of interest to highway administrators and decision makers at all levels of state government.

NORTH CAROLINA HIGHWAY STATISTICS AND COST FINDINGS

Mileage and Travel

Nearly 44 billion VMT occurred on 92,719 mi of North Carolina highways in 1984:

- Fifty-eight percent of all travel occurred in rural areas.
- Forty-two percent of the travel occurred in urban areas that contained only 19 percent of the total highway system mileage.
- More than 14 percent of travel was on the Interstate system.
- The arterial system (including the Interstate) constitutes less than 10 percent of the highway mileage but carried more than 60 percent of the travel.
- The collector system represents 23 percent of the public road mileage and carried 26 percent of the travel.
- The local functional class system represents 68 percent of the total highway mileage but carried only 14 percent of the total travel.

TABLE 5 SUMMARY OF COMBINED COSTS AND UNIT COSTS FOR 1993 (ZERO FUNDING SCENARIO)

	Combined Costs ($ millions)				Unit Costs per 1,000 VMT			
	Operating	Accident	Time	Total	Operating	Accident	Time	Total
Rural								
Interstates	2,656.225	75.811	1,269.601	4,001.637	371.50	10.60	177.57	559.67
Other principal arterials	1,935.560	143.408	1,089.538	3,168.506	291.50	21.59	164.09	477.18
Minor arterials	1,179.174	99.551	754.940	2,033.665	285.10	24.07	182.53	491.7
Major collectors	2,888.817	283.078	1,973.052	5,144.946	273.20	26.77	186.59	486.56
Minor collectors	685.078	79.727	577.654	1,342.459	243.80	28.37	205.57	477.74
Urban								
Interstates	1,459.753	66.983	818.583	2,345.319	360.70	16.55	202.27	579.52
Other freeways and expressways	712.517	46.934	415.260	1,174.711	263.70	17.37	153.69	434.76
Other principal arterials	2,972.785	293.071	3,536.070	6,801.926	295.30	29.11	351.25	675.66
Minor arterials	1,336.634	147.071	1,433.745	2,917.45	262.60	28.59	281.68	573.17
Collectors	360.233	25.507	315.731	701.471	335.10	23.73	293.70	652.53
Total	16,186.776	1,261.143	12,184.177	29,632.096	298.15	23.23	224.42	545.80

TABLE 6 SUMMARY OF COMBINED COSTS AND UNIT COSTS FOR 1993 (MAINTAIN-1984-CONDITIONS SCENARIO)

	Combined Costs ($ millions)				Unit Costs per 1,000 VMT			
	Operating	Accident	Time	Total	Operating	Accident	Time	Total
Rural								
Interstates	1,996.280	75.811	1,051.596	3,123.687	279.20	10.60	147.08	436.88
Other principal arterials	1,760.260	135.802	984.000	2,881.000	265.10	20.45	148.33	433.88
Minor arterials	1,084.050	99.069	674.637	1,857.756	262.10	23.95	163.11	449.16
Major collectors	2,768.270	282.096	1,896.932	4,947.298	261.80	26.68	179.40	467.88
Minor collectors	637.030	79.727	543.883	1,260.640	226.70	28.37	193.55	448.62
Urban								
Interstates	1,085.000	64.092	688.792	1,837.884	268.10	15.84	170.20	454.14
Other freeways and expressways	617.680	44.304	366.330	1,028.314	228.60	16.40	135.58	380.58
Other principal arterials	2,569.100	293.071	3,290.710	6,152.881	255.20	29.11	326.88	611.19
Minor arterials	1,171.210	146.599	1,362.939	2,680.748	230.10	28.80	267.77	526.67
Collectors	308.200	25.511	289.877	623.588	286.70	23.73	269.65	580.08
Total	13,997.080	1,246.082	11,150.714	26,393.876	257.82	22.95	205.39	486.16

TABLE 7 SUMMARY OF COMBINED COSTS AND UNIT COSTS FOR 1993 (100 PERCENT FUNDING SCENARIO)

	Combined Costs ($ millions)				Unit Costs per 1,000 VMT			
	Operating	Accident	Time	Total	Operating	Accident	Time	Total
Rural								
Interstates	1,827.540	75.811	1,006.339	2,909.690	255.60	10.60	140.75	406.95
Other principal arterials	1,553.760	116.861	867.664	2,538.285	234.00	17.60	130.67	382.27
Minor arterials	961.620	96.421	606.202	1,664.243	232.50	23.31	146.57	402.38
Major collectors	2,417.216	277.672	1,730.055	4,424.943	228.60	26.26	163.61	418.47
Minor collectors	543.735	79.727	495.848	1,119.310	193.50	28.37	176.46	398.33
Urban								
Interstates	974.113	63.995	658.881	1,696.989	240.70	15.81	162.81	419.32
Other freeways and expressways	564.718	44.304	351.663	960.685	209.00	16.40	130.15	355.55
Other principal arterials	2,383.866	292.991	3,226.910	5,903.767	236.80	29.10	320.54	586.44
Minor arterials	1,161.029	146.599	1,357.028	2,664.656	228.10	28.80	266.61	523.51
Collectors	250.153	25.511	261.735	537.399	232.70	23.73	243.47	499.90
Total	12,637.750	1,299.894	10,562.327	24,419.971	232.78	22.97	194.55	449.80

Service

Forty-four miles of the rural highway system were operating at traffic speeds of less than 30 mph during peak hours.

Thirty-nine miles of the urban freeway and expressway system were operating at traffic speeds of less than 35 mph during peak periods.

Pavement Condition

Four thousand eight hundred fifty-seven (4,857) miles of the rural and 1,187 mi of the urban arterial and collector road system has deteriorated to a poor or very poor condition (i.e., PSR ≤ 2.5). Local roads were not sampled, and the deteriorated mileage on this system is not known.

Safety

The respective accident rates for fatal and nonfatal injuries were 5.1 and 66 per 100 million VMT in rural areas.

The respective accident rates for fatal and nonfatal injuries were 3.0 and 143 per 100 million VMT in urban areas.

Capacity

The rural arterial and collector highway system had 89 mi operating with traffic congestion during peak periods.

The urban arterial and collector highway systems had 47 mi operating with traffic congestion during peak periods.

Deficiencies

A summary of data on pavement, operational, and geometric deficiencies is given in Table 8.

TABLE 8 PAVEMENT, OPERATIONAL, AND GEOMETRIC DEFICIENCIES

Functional System	Miles Deteriorated	Percentage of System
Pavement Deficiency (may include combinations of deficiencies)		
Rural		
Interstates	65	11
Other principal arterials	525	26
Minor arterials	383	19
Major collectors	809	8
Minor collectors	842	9
Urban		
Interstates	60	30
Other freeways and expressways	74	35
Other principal arterials	446	27
Minor arterials	307	14
Collectors	285	21
Operational Deficiency (may include combinations of deficiencies)		
Rural		
Interstates	34	6
Other principal arterials	174	9
Minor arterials	40	2
Major collectors	107	1
Minor collectors	0	0
Urban		
Interstates	26	13
Other freeways and expressways	27	13
Other principal arterials	NA	NA
Minor arterials	NA	NA
Collectors	NA	NA
Geometric Deficiency (may include combinations of deficiencies)		
Rural		
Interstates	0	0
Other principal arterials	503	25
Minor arterials	652	33
Major collectors	1,299	12
Minor collectors	1,065	12
Urban		
Interstates	NA	NA
Other freeways and expressways	NA	NA
Other principal arterials	NA	NA
Minor arterials	NA	NA
Collectors	NA	NA

Highway Costs

The cost to eliminate the current (1984) backlog of needed improvements and to fund the expected ongoing needs through 1993 is $5.2 billion for rural areas and $1.7 billion for urban areas (Figures 1 and 2).

Annual cost required through 1993 to fund the full-needs scenario is $694.4 million.

Annual cost required through 1993 to fund the maintain-1984-conditions scenario is $408.9 million.

The total annual cost of travel in 1993 including operating, accident, and time components under the zero funding scenario is $29.6 billion. The unit operating cost for this funding scheme is $545.80 per 1,000 VMT (Table 5).

The total annual cost of travel in 1993 including operating, accident, and time components for the maintain-1984-conditions funding scenario is $26.4 billion. The unit operating cost for this funding scheme is $486.16 per 1,000 VMT (Table 6).

The total annual cost of travel in 1993 including operating, accident, and time components for the full-needs scenario is $24.4 billion. The unit operating cost for this funding scheme is $449.80 (Table 7).

The annual cost difference between the full-needs and the maintain-1984-conditions funding scenarios is $285.5 million. The cost of travel difference between these scenarios is $2.0 billion for 1993.

CONCLUSIONS

It is recognized that the North Carolina Department of Transportation's funding capability for construction, reconstruction, resurfacing, restoration, rehabilitation, and normal maintenance of highways is variable and depends on several political, social, and economic factors. It is further recognized that this funding capability is quite sensitive to and can be affected by changes in the level, character, or distribution of funds by local government, or all three. However, given the study findings, it can be concluded that the anticipated funding levels through 1993 will not be adequate to meet ongoing needs or eliminate the backlog of deficiencies on the existing highway system in North Carolina. It can also be concluded that the total 1993 operating cost on that system will be more than $27 billion.

That operating cost could be reduced by approximately $3 billion under the full-needs funding scenario.

A final study conclusion is that the HPMS procedures will continue to be an important tool for identifying highway improvements that should lead to optimal achievement of North Carolina's highway transportation goals.

REFERENCES

1. *North Carolina Highway Needs—A Summary.* Planning and Research Branch, North Carolina Department of Transportation, Raleigh, Aug. 1977.
2. *Highway Performance Monitoring System Field Manual for the Continuing Analytical and Statistical Data Base.* FHWA, U.S. Department of Transportation, Jan. 1984.
3. *Highway Performance Monitoring System Analytical Process,* Vol. 1, Version 2.0: *Executive Summary.* FHWA, U.S. Department of Transportation, March 1983.
4. *Highway Performance Monitoring System Analytical Process,* Vol. 2, Version 2.0: *Technical Manual.* FHWA, U.S. Department of Transportation, Jan. 1986.
5. *Highway Performance Monitoring System Analytical Process,* Vol. 3, Version 2.0: *User's Guide.* FHWA, U.S. Department of Transportation, Jan. 1986.
6. *Highway Performance Monitoring System Analytical Process,* Vol. 4: *System Documentation.* FHWA, U.S. Department of Transportation, July 1983.
7. *Highway Performance Monitoring System Analytical Process: Spreadsheet/Graphics Applications.* FHWA, U.S. Department of Transportation, Jan. 1986.
8. *Highway Functional Classification: A Management Tool.* FHWA, U.S. Department of Transportation, Nov. 1982.
9. Secretary of Transportation. *The Status of the Nation's Highways: Conditions and Performance.* Report to the 99th Congress, June 1985.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.

Sufficiency Ratings for Secondary Roads: An Aid for Allocation of Funds?

CLETUS R. MERCIER AND JAMES W. STONER

A new model that can be used to make sufficiency ratings of secondary roads is described briefly. The calibration and scales used with the rating criteria are described in some detail, with particular emphasis on both the linear and the nonlinear features of the scales. The flexibility of the model, particularly with the use of variable standards, is noted. This feature makes it possible to prepare new priority rankings based on revised functional classification or design standards, or both, for local use. It is also suggested that the model could be used as an aid in developing new road use tax allocation formulas. It is concluded that, though many other factors need to be considered, use of the same model across jurisdictions would provide realistic statewide needs assessments.

State highway organizations commonly use a numerical evaluation system for priority planning of roadway improvements. Most evaluation systems in use are patterned after a numerical rating scheme, first developed by the Arizona Highway Department in 1946 (*1–3*), that describes a highway's "sufficiency." The sufficiency rating method assigns a point score to each section of road on the basis of its actual condition and its ability (or inability) to carry its traffic load in a safe and efficient manner. There have been attempts to develop a successor to sufficiency rating, designed to take advantage of the computer's speed and flexibility by including additional factors and a sophisticated calibration procedure. However, most of the successors have failed to gain wide acceptance, although a formalized pavement management system (PMS) is gaining prominence (*4*).

The sufficiency rating for a given segment of road is a composite score; it represents the sum of evaluation scores of a number of highway and traffic elements. Much commonality exists among the lists of rating elements or criteria used by the various states, but there are also differences.

The differences are of two kinds. There are differences in choice of criteria used to evaluate a given road segment. There is a list of commonly used criteria, but some divergence in the choice of those actually used. There are also differences in how the criteria are actually weighted. Both produce some variations in rating formulas.

These differences can be explained in two ways. First, there are differences in conditions among the states. Second, there are valid differences in opinion, especially in the perception of relative importance (*5*). Sufficiency rating is often described as empirical, or based on practical experience (*1*).

C. R. Mercier, College of Engineering, Iowa State University, Ames, Iowa 50010. J. W. Stoner, College of Engineering, The University of Iowa, Iowa City, Iowa 52240.

Although most state highway organizations use sufficiency rating systems for priority planning, the practice is not prevalent among county highway organizations. There have been sufficiency rating systems developed and used by some county highway organizations; however, none appears to have survived the tenure of the administrator who put it in place or to have been used in other jurisdictions.

State and county highway officials face a similar problem in decision making: how should available funds be allocated? For state officials, an evaluation system that makes it possible to measure a given road segment's sufficiency relative to road segments in other parts of the state is important. A system that makes rational comparisons of needs greatly simplifies problems in priority programming. Used efficiently, it leads logically to allocation of available funds within a jurisdiction.

Another important consideration is the wide variation in average daily traffic (ADT) over the various primary roads. Without some "leveler" it would be nearly impossible to choose between taking care of the needs of a limited-access, four-lane highway with large ADT and those of a two-lane state highway carrying considerably less traffic. Therefore both functional classification and applicable design standards play key roles in priority programming. In addition, priority decisions can be more easily defended when an evaluation system is used, but the large size of a state road network would make specific challenges less likely.

In contrast, officials responsible for priority programming for secondary roads (particularly at the county level) are commonly quite familiar with all segments of the road network, and priority decisions often are made informally. Also, there are fewer variations in the road characteristics, as defined either by functional classification or by differences in design standards. As an example, area service roads make up about two-thirds of the secondary road network in Iowa, and most have a gravel surface, but it would be hard to find any significant difference in the design characteristics of most of these roads, no matter what the traffic volume. Therefore it is difficult to differentiate between two roads carrying different traffic loads in preparing priority lists. County officials also find it more difficult to defend their programming decisions because of the informal nature of the decision-making process. Therefore there is some interest in a sufficiency rating system for secondary roads.

In 1985 a research project, sponsored by the Iowa Department of Transportation (Iowa DOT), was completed that resulted in a model that could be used as a sufficiency rating system for secondary roads. The model is also empirically based—on the Arizona format and the experience of local engineering practitioners (6).

The resulting system has 14 rating elements that represent the expressed preferences of local practitioners. The preferences were expressed through a survey of county engineers in Iowa. Relative weights were also assigned on the basis of the opinions divulged in the same survey. The development of the scales for these rating elements and how they can affect the allocation of highway funds are described in this paper.

MODEL

Fourteen rating elements were selected for use with the proposed sufficiency rating system. They have been organized into three categories, similar to the Arizona format, and assigned relative weights. Table 1 gives the proposed list of rating elements and their suggested weights.

Form of Model

The basic model for the sufficiency rating system is a simple linear additive model. The maximum possible scores for the selected rating elements were determined from analysis of the data received. What remained was to solve the problem of how to assign scores when the rated road segment fails to meet the expected standard for a given rating element. To do this requires answers to two questions:

1. What is a defensible set of standards that could be applied to the rating elements selected?
2. Is there a scaling calibration that can be used with each rating element and that would yield meaningful scores when the rated road segment failed to meet the desired standard?

The answers to these two questions are critical to the problem of the assignment of scores. In the next two subsections the issues raised by these questions will be addressed and appropriate answers will be suggested.

Standards for Rating Elements

The determination of appropriate standards to apply to the rating elements is intermixed with economic and social issues: what level of financial commitment is the public willing to make to build and maintain the transportation infrastructure, and what is the dollar value of personal comfort, pain and suffering (due to traffic injury), and human life (when a person is killed in a traffic accident)?

Though these issues will probably never be settled, engineering practitioners have adopted standards that are reasonably consistent with prevailing public opinion. Evidence of public opinion is provided by the level of funding legislative bodies have allocated and by the force of individual and group pressures.

The resulting design standards are adopted by highway agencies for use on all of the different classes of roads throughout their jurisdictions. Comparable sets of design standards have been adopted by many state highway organizations. These standards are similar in many respects, but they also reflect local conditions. Design standards represent prevailing professional opinion on appropriate standards or norms for building a given road to serve expected traffic needs.

Iowa has developed design guides that call for higher standards of construction for roads carrying heavier volumes of traffic (and costing more) and concomitant lower standards for roads carrying less traffic. These design guides were developed by Iowa DOT staff for the 1982–2001 Quadrennial Needs Study (7) in consultation with the State Functional Classification Review Board (as specified by law), members of the County Engineers Association, and the League of Iowa Municipalities.

Because the design guides are prepared in consultation with so many interested parties (there are several nonengineers on the Review Board), there is an inference that the lowered standards are acceptable to the public. Further, it would appear logical that there is little reason to exceed the lowered standards for the lightly traveled roads, except when it can be done at little extra cost. Similarly, a rational approach to evaluation of sufficiency—a comparison with established ideals—should be based on current design standards for that road classification.

Therefore the sufficiency rating model developed for secondary roads incorporates applicable design standards from the

TABLE 1 FINAL PROPOSED SUFFICIENCY RATING SYSTEM MODEL

Rating Category	Item Rated	Maximum Points
Condition and maintenance experience (35 points)	Foundation	9
	Wearing surface	9
	Drainage	8
	Maintenance economy	9
Safety (40 points)	Accident rate	6
	Hazards	9
	Stopping sight distance	8
	Passing sight distance	5
	Traffic control	6
	Horizontal alignment	6
Service (25 points)	Pavement (roadbed) width	9
	Ride quality	5
	Snow problems	6
	Surface type (unpaved)	5
	Shoulder width (paved)	(5)

alternate design guide developed by Iowa DOT staff for the Needs Study. This guide was chosen for the model even though many counties use the Farm to Market Design Guides. It was chosen because of its breakdown of area service roads into three categories based on ADT. This provides for lower standards for lightly traveled area service roads. It also represents what are expected to be the design standards of the future.

Failure of a rated road segment to meet the applicable standard would cause a lowered score for that rating element. Established ideals for rating elements not covered by a design standard are based on current practices as evidenced by a combination of standards used in other sufficiency rating systems and on local practices.

Scaling Factors

An assessment of the maximum point value for a given rating element is made when the road segment meets or exceeds the current standard. However, a given rated road segment will sometimes fail to meet the current standard for each of the rating elements, making it necessary to develop some sort of scale to describe how close it comes to meeting that standard. Maximum point values for each of the rating elements are given in Table 1, so what is needed is a set of graduated scales for each.

Existing systems use, for the most part, a sequence of point values that is approximately linear in character. In most instances there is a score (often at about the middle of the scale) that represents an average value, below which a road segment is considered intolerable. The concept of tolerability is based on the supposition that, for each rating element, there is a tolerable standard that is less desirable than the ideal but that is still considered safe, or at least provides good service. The tolerable point is the lowest point on the scale permissible under current highway transportation requirements. Below that level, the rated road segment is considered intolerable in terms of that rating element.

The Iowa DOT uses a sufficiency rating system to evaluate Iowa's primary roads. The calibration system used establishes tolerable levels for each rating element of the system. (In this context, the term "scale" is used to describe the graduations along an axis, and calibration is the numerical values assigned to the graduations.) In each instance, the tolerable point is half of the maximum point value, rounded down to the next digit when the maximum point value is not an even number.

This general calibration method is used for the secondary road model (with slight variations), graduated linearly with decreasing values below the maximum score. Accompanying statements have used descriptors of "excellent," "good," "fair" (at tolerable scales), and "poor," together with status descriptions for each score. A summary of the model's scoring method is given in the next section.

However, there are some rating elements in the model that do not lend themselves well to the linear scale concept. They include elements grouped under the category of Safety. The score represents an accumulation of potential safety risks or hazards occurring along the rated road segment. Their existence represents a possible safety hazard, or deficiency, and they tend to be site specific instead of occurring regularly along the road. The rating elements are the type that can be counted (two narrow bridges are more hazardous than one). Some are based on design standards for secondary roads. An example is "narrow structures"—structures narrower than 20 ft (6 m). Any structure less than 20 ft wide is assumed to represent a safety hazard.

This suggests that part of the score for a rated road segment under the category of Safety could be based on the results of an evaluation of its relative safety. Deductions from a maximum value would be made for "conditions that exist on the road segment that constitute a possible threat to safe operation of the motor vehicle on that road."

Under this system, deficiency points would be assessed for the existence of a list of "threats to safe driving," using a predetermined point deduction for each deficiency. Road segments of varying length would be made comparable by adjusting for length. There would be no negative scores for safety, but a given road segment could receive a zero score.

The scaling system developed for use with the model is briefly described in the next section. The set of scales is described completely in Volume 2 of the project report (*8*).

Rating Scale Calibration

The rating scale system was designed to provide relative scores for the sufficiency ratings for each road in order to place road improvement projects in some priority order. In theory there is no score that "fails," but the rating system allows for comparison of scores for roads that carry different amounts of traffic. The system is neutral when comparing roads in different parts of the county highway department's jurisdiction, and it recognizes the differences in needs of the more heavily used arterial roads and the area service roads that carry substantially less traffic. This means that should two roads have scores of 70 and 65, the one with the lower score should have a higher priority for improvement, even if the road with the lower score carries substantially less traffic.

The validity of comparison of the scores of roads of different classifications is assured by including the variations in design standards for each road classification. This affected rating standards for several criteria under the categories of Safety and Service.

The scale calibration described next represents a sampling of the scales provided for each criterion used in the model and included in the project report.

Linear Scales

The format used in describing the calibration for the criteria in the category of Condition and Maintenance Experience is rather consistent. It provides a brief description of each criterion followed by the range of possible scores and descriptors designed to aid the evaluator in determining an appropriate score. An example is the calibration for Road Foundation, taken from the project report:

- Foundation—evaluated by considering adequacy of drainage ditches, breakup of surface, nonuniform settlement and lateral support, and condition of foreslopes. Maximum score = 9.
- Excellent, 8–9. No evidence of base failure. Foreslopes in excellent condition.

- Good, 6–7. Occasional evidence of minor base failure, fully correctable by spot repairs. No need for extensive reworking.
- Fair, 5. Frequent base failure requiring heavy maintenance. Causes reduction in traffic speeds below design speed. Should be considered for reconstruction. "Tolerable."
- Poor, 1–4. Severe base failure throughout rated section, extreme "washboard" condition. Traffic speeds substantially reduced. Reconstruction necessary.

Nonlinear Scales

Scores for some of the rating elements under the category of Safety are derived somewhat differently. As noted earlier, there are some rating elements that do not lend themselves to the linear scale used for the category of Condition and Maintenance Experience. Instead, deficiency points are assessed for the existence of "threats to safe driving." A predetermined point deduction is charged for each deficiency.

Two brief examples, using the rating criteria of Accident Rate and of Hazards, will help explain the concept. Each formula provides for comparison of road segments of varying lengths by use of the factor L, length of rated segment in miles.

- Accident Rate: Deficiency points are assessed for each accident occurring on that road segment over the past 5 years. Relative weights of each accident vary according to severity of the accident. Property damage accidents result in one deficiency point, while personal injury accidents are four and a fatal accident would be twelve deficiency points. The score for a given road segment uses the formula

$$\text{Rating} = 6 - (N/L)$$

where N is the sum of all deficiency points and L is the length of the rated road segment in miles. The maximum score is 6, but the minimum score would be 0.

- Hazards: Deficiency points are assessed for each hazard not included in any other rating element. These hazards include

1. Narrow structures (less than 20 ft),
2. Structure with poor approach alignment,
3. Railroad crossing at grade without automatic signals,
4. Abrupt or severe grade changes, and
5. Other fixed structures extending into the traveled way.

Rating scores are based on the average number of hazards per mile of roadway using the formula

$$\text{Rating} = 9 - 2(N/L)$$

where N is the number of hazards encountered and L is the length of the rated road segment in miles. The number 2 represents the perceived weighted severity index of the effect of the hazards on driving safety. The maximum score is 9 and the minimum is 0.

The effect of these factors (Accidents and Hazards) on the overall score for a road segment can be varied according to the perceived importance of the factors to driving safety. The weight of the type of accident can vary as well as the weighted severity index for hazards. Indeed, the index can vary according to the type of hazard encountered. Engineering practitioners can vary these weights—using the same rationale that was used to promulgate design standards. The basis for choice of relative weights could be "prevailing professional opinion" arising from the force of public opinion.

A third general type of scale, also nonlinear, is represented by some of the rating criteria under the category of Service as well as under Safety. This type of scale uses the applicable design standard for the rated road segment according to the adopted Design Guide. An example is the criterion of Pavement (roadbed) Width referring to traveled way for an unpaved road. (Shoulder width for paved roads is covered under a separate criterion.) The calibrations, taken from the project report, are

- Pavement (roadbed) Width—used to reflect inadequate traveled way widths as determined by a comparison with the appropriate design standard. Maximum score = 9.
- Excellent, 8–9. Width of pavement or traveled way meets or exceeds the width specified in the appropriate design standard.
- Good, 6–7. Width of pavement or traveled way is not more than 2 ft (0.6 m) less than the design standard.
- Fair, 5. A tolerable width. Width of pavement or traveled way is 2 ft (0.6 m) to 4 ft (1.2 m) less than the design standard.
- Poor, 1–4. Not tolerable. Needs to be wider. Width falls short of the design standard by at least 4 ft (1.2 m).

INCENTIVES FOR CHANGE

The exercise just described is potentially useful as an internal guide in programming for secondary road improvements, particularly if the variable design standards are used. An individual county can use it to provide the basis for allocation of funds for secondary road needs.

Properly used, the rating system provides a ranking for developing annual programs, and it can also be used to assist in maximizing stated objectives. However, it has the potential for use on a larger scale in the allocation of road use tax funds on the statewide level for highway construction, rehabilitation, and maintenance. State road use tax funds are commonly split among several jurisdictions, and there is never enough to go around. Currently, the combination of a deteriorating physical plant for primary highways with heavier traffic loads and heavier axle loads for trucks has caused the initiation of new discussions on the allocation of highway funds to the various political jurisdictions.

In rural states, rural-dominated legislatures have generally been able to retain a significant proportion of the road use tax funds for secondary roads, and will probably be able to do so for some time. However, the recent problems plaguing the farm economy have accelerated the displacement of farmers from rural to urban areas, further reducing an ever-shrinking farm population. Future reapportionment of state legislatures is likely to produce a more urban-oriented body of lawmakers, one less favorable to rural issues.

A problem that will face legislatures in the future, whatever their makeup, will be how to significantly increase the total highway funds available to a given political jurisdiction.

Though the dollar amount of funds is not really a fixed amount, it should be realized that there have been several significant increases in the gasoline tax during the past few years by both the federal government and state governments. At the time of writing, gasoline taxes in Iowa total $0.25 (out of about $1.00 per gallon at the pump), $0.16 for the state and $0.09 for the federal government. This is up from $0.07 for the state in 1978 and $0.04 for the federal government as recently as 1983. Any effort to increase that amount in the near future is likely to meet some resistance, at least at the state level.

Therefore any significant increase in the highway funds any political jurisdiction receives from road use tax funds is not likely to be accomplished by additional taxes but by a change in the allocation of funds. And because allocation is based at least to some extent on the results of needs studies, much time is likely to be spent analyzing the results of needs assessment studies.

Secondary roads have traditionally fared well when needs have been considered, given past methods of determining needs. It is difficult to ignore the results of needs studies, but it is possible to redefine needs. One way to redefine needs is to reconsider some of the currently accepted standards. For example, should a bridge that is 16 to 18 ft wide be considered inadequate (intolerable) for a road carrying fewer than 50 vehicles per day, even if it is considered safe, can carry expected loads, and is properly signed?

Design and Classification Changes

What is needed is a close reexamination of the way secondary roads are used. Some do carry moderate to heavy traffic loads, but many serve only as an access to abutting property. As farms continue to increase in size [in Iowa, from 276 acres (108 ha) in 1976 to an estimated 303 acres (122 ha) in 1986] and fewer farmsteads are occupied, there will be an increasing number of roads that merely serve as access to farmland.

Attempts have been made in the past to abandon some of these roads with title reverting back to the owners of adjoining property. However, many of these actions have encountered considerable resistance, making the total miles of secondary roads abandoned in most jurisdictions fairly insignificant. A recent study by Baumel et al. (*9*) generally supports these property owners. The study indicates that some of the roads normally considered candidates for abandonment should be retained in the road network because the benefits to the public of keeping the roads open equal or exceed the costs of closing the roads. A major factor in the analysis was the higher travel cost of farm equipment such as tractors, wagons, and combines.

The significance of farm equipment travel costs in the study suggests that the very-low-volume roads kept open need not meet the same standards as collector roads. This is because of the low speed at which much of this equipment moves. The Baumel report did suggest that some groups of low-volume roads could be converted to private roads, with the landowners assuming the maintenance costs. If the local owners perform the maintenance, these are likely to become "minimum maintenance" roads. Likewise, little reconstruction is likely to be done.

Changes in the characteristics of the use of these very-low-volume roads have been recognized by county officials in Iowa, as well as several other states, by their designation of some public roads as minimum maintenance roads. This represents the creation of a new functional classification carved out of the area service functional class. To date Iowa's experience with this concept has been good, and, by mid-1986, 45 percent of the counties had voluntarily adopted the classification and had designated about 10 percent of their secondary road mileage as minimum service (Service B classification) (*10*). By mid-1987, this was up to about 80 percent. Indications are also that this percentage will increase slightly over time. (Though Baumel et al. did refer to this possible approach, it was not examined in the report analysis.)

Iowa has also taken steps to provide design guides that provide for a wider range of design speeds, from 30 to 55 mph. The 1985 guidelines are given in Table 2. (Road surfacing is not covered by the design guide because that issue is addressed by a separate policy.) The Guide meets the design criteria set out in the current edition of the AASHTO *A Policy on Geometric Design of Highways and Streets*, Chapter V, and meets the expectations of courts of law with regard to tort liability (*11*). The new design standards also should manage to achieve cost savings through use of the varying design speeds.

It would appear appropriate, therefore, to take a closer look at the way secondary roads are used and to consider changes in design standards and possibly even functional classification. This needs to be done on a national basis because designation of functional classification is so closely tied to the disbursement of Federal-Aid Secondary funds. Without this, some of the needed reclassification would not occur. After all, who would expect county officials to downgrade a road's functional classification if it meant the loss of revenue, even if the lower classification were clearly warranted?

New Needs Assessment

Adoption of significant changes in design standards or functional class, or both, would require a new needs assessment for each jurisdiction because many existing roads would change. Should this be done, a reasonably uniformly applied sufficiency rating system could yield results that would better define secondary road needs statewide. If all jurisdictions used the same bases for needs assessment, a summation of the needs, both short and long range, could be used to aid in the determination of an allocation formula.

It should be noted here that a new needs assessment is not likely to be the only concern in determining a new allocation formula, though it would probably be an important factor. County governments in Iowa already are facing significant losses in revenue for use on secondary roads from two current sources. One is property tax revenue that is decreasing because of the lower land values throughout the state, and the other is loss of federal funds—revenue sharing and other federal sources. (Counties have consistently used a portion of their revenue sharing funds for secondary roads.)

TABLE 2 DESIGN (AASHTO) GUIDELINES FOR RURAL COLLECTORS, 1985

DESIGN ELEMENTS		ALL ROADWAYS														
(1) ADT--DHV								100-200			200-400			Over 400		
--Design Year (in 20 yrs.)		0-600			600-750			750-1500			1500-3000			Over 3000		
--Current Year		0-400			Over 400			500-1000			1000-2000			Over 2000		
TERRAIN		Flat	Rolling	Mount	Flat	Rolling	Mount	Flat	Rolling	Mount	Flat	Rolling	Mount	Flat	Rolling	Mount
DESIGN SPEED	mph	40	30	20	50	40	30	50	40	30	60	50	40	60	50	40
STOPPING SIGHT DISTANCE	ft	275-325	200	125	400-475	275-325	200	400-475	275-325	200	525-650	400-475	275-325	525-650	400-475	275-325
MAXIMUM CURVATURE	degrees	12.25	22.75	53.5	7.5	12.25	22.75	7.5	12.25	22.75	4.75	7.5	12.25	4.75	7.5	12.25
(2) MAXIMUM GRADIENT	%	7	9	12	6	8	10	6	8	10	5	7	10	5	7	10
PAVEMENT/SURFACING WIDTH	ft	20	20	20	22	22	20	22	22	20	24	24	22	24	24	24
SHOULDER WIDTH	ft	2	2	2	4	4	4	6	6	6	8	8	8	8	8	8
ROADWAY TOP WIDTH	ft	24	24	24	30	30	28	34	34	34	40	40	38	40	40	40
(3) BRIDGE WIDTH--NEW	ft	24	24	24	28	28	26	28	28	26	32	32	30	40	40	40
(4) BRIDGE WIDTH--EXISTING	ft	22	22	22	22	22	22	22	22	22	24	24	24	28	28	28
FORESLOPE		3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1	3:1
(5) CLEAR ZONE		10	10	10	BG	10	10	BG	10	10	BG	BG	10	BG	BG	10

NOTES:

(1) DHV governs
(2) Maximum Gradient may be steepened by one percent (1%) for short distance--(less than 500')
(3) a. Bridges over 100 feet long and DHV over 200, width may be traveled way plus three feet (3') each side
b. Design Loading should be HS-20
(4) a. For bridges less than 100 feet in length, over 100 feet analyze individually
b. Design Loading should be HW-15
c. Existing bridge width is considered to be at least pavement width
(5) CLEAR ZONE = 10' for 40 mph and below and according to Barrier Guide (BG) for 50 mph and above (Clear Zone Table in I.M. 3.215)

SUMMARY AND CONCLUSIONS

The objective of this study was to produce a sufficiency rating system that could be used to evaluate the adequacy of secondary roads in Iowa. The system developed should be reasonably easy to use yet yield results that are compatible with processes currently used in priority planning.

The final sufficiency rating formulation appears to do this, plus it provides a possible bonus. It could also be used as a basis for statewide allocation of funds, particularly if a more variable set of standards were used. The suggested lower standards for the very-low-volume roads can be factored into the model by using the different design guides and the calibrations employed with the rating criteria. Though a more variable set of standards would make the evaluation process slightly more complicated, it would provide more realistic rating scores that would more accurately reflect the nature of traffic on and the frequency of use of a given road.

In addition, use of the same set of scales across jurisdictions would make it possible for lawmakers to make more realistic needs assessments. More uniform guidelines can be used for needs assessment in the jurisdictions responsible for secondary roads. Then, comparison can be made between those needs and those of primary roads or urban streets, or both, using similar methods of needs assessment. Like jurisdictions can then be treated more equally in the allocation process.

The allocation process is still, of course, political, but lawmakers are provided with better information about needs. Many states already use either a sufficiency rating system or a similar numerical evaluation process to determine primary road needs more objectively. The same can be done with secondary roads to provide lawmakers with better data for use in determining allocation formulas for road use tax funds.

REFERENCES

1. K. Moscowitz. Numerical Ratings for Highway Sections as a Basis for Construction Programs. *Bulletin 17*, HRB, National Research Council, Washington, D.C., 1948, pp. 28–39.
2. W. E. Willey. Arizona's Experience with Sufficiency Ratings. *Bulletin 53*, HRB, National Research Council, Washington, D.C., 1952, pp. 3–7.
3. C. V. Zegeer and R. L. Rizenbergs. Priority Programming for Highway Reconstruction. In *Transportation Research Record 698*, TRB, National Research Council, Washington, D.C., 1979, pp. 15–23.
4. T. F. Humphrey. *NCHRP Synthesis of Highway Practice 84: Evaluation Criteria and Priority Setting for State Highway Programs.* TRB, National Research Council, Washington, D.C., Nov. 1981, 32 pp.
5. J. A. Swanson. General Comments on Sufficiency Rating Procedures. *Bulletin 53*, HRB, National Research Council, Washington, D.C., 1952, pp. 7–11.
6. C. R. Mercier. Sufficiency Rating for Secondary Roads: Model Development. In *Transportation Research Record 1076*, TRB, National Research Council, Washington, D.C., 1986, pp. 7–13.
7. R. L. Carstens. *Quadrennial Needs Study, 1982–2001.* Iowa Department of Transportation, Ames, 1982.
8. C. R. Mercier. *A Sufficiency Rating for Secondary Roads in Iowa.* Project Report HR-264. Iowa Department of Transportation, Ames, 1985.
9. C. P. Baumel, C. A. Hamlett, and G. R. Pautsch. *The Economics of Reducing the County Road Systems: Three Case Studies in Iowa.* Project Report HR-242. Iowa Department of Transportation, Ames, 1987.
10. C. R. Mercier and L. E. Richardson. Area Service B Classification: A Legal Approach to Setting Road Maintenance Levels for Little Used Low-Volume Roads. In *Transportation Research Record 1106*, TRB, National Research Council, Washington, D.C., 1987, Vol. 1, pp. 34–38.
11. R. L. Carstens. Highway-Related Tort Claims to Iowa Counties. In *Transportation Research Record 833*, TRB, National Research Council, Washington, D.C., 1981, pp. 18–24.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.

Solving the Suburban Mobility Problem: Two Case Studies in the Application of Collaborative Problem-Solving Techniques

LARRY L. BYE, FRANCES A. COOPER, AND JAMES R. LIGHTBODY

In this paper are reviewed two transportation planning projects in which major stakeholders from both the public and the private sectors participated through a collaborative-planning or consensus-building process. Both projects took place in Santa Clara County, California, site of the burgeoning Silicon Valley. The first project, Transportation 2000, resulted in the adoption of a midrange transportation plan in a policy environment characterized by 15 separate municipal jurisdictions with a history of competition for the fiscal benefits of development and without a formal mechanism for coordination of transportation and land use policies. The second project, the Fremont–South Bay Alternatives Analysis, is taking place in a similarly complex environment. Its goal is to choose a locally preferred transportation alternative for the corridor by May 1988. The collaborative planning strategies that are assisting stakeholders in reaching agreement and making commitments to implementation are outlined. These strategies include identifying stakeholders, one-on-one interviews, information management, facilitated small-group sessions, working with the press, involving rank-and-file community members, and quantitative public opinion surveys. The paper includes a critique of both projects as a guide to others who may want to undertake such efforts.

Traffic congestion threatens the business climate and quality of life of most rapidly growing metropolitan areas in the United States and is likely to get worse. Expansion of roadway capacity cannot solve the problem. Because of topography, environmental concerns, and other factors, not enough additional lanes on freeways can be built to meet demand.

Other approaches to capacity expansion are needed: transitways and high-occupancy-vehicle lanes, rail systems, and bus system improvements. However, there are problems with transit underutilization and lack of public support for expensive transit-related capital investments.

Demand management is also required: major employers must share the responsibility for reducing solo automobile use. Parking policies must be reexamined. Yet many employers and key participants in land development resist these changes.

L. L. Bye, Communication Technologies, 260 California Street, Suite 803, San Francisco, Calif. 94111. F. A. Cooper, Department of Public Administration, California State University, Hayward, Calif. 94542. J. R. Lightbody, Santa Clara County Transportation Agency, 1570 Old Oakland Road, Suite 202, San Jose, Calif. 95131.

The congestion problem originates in the sheer pace of job creation and residential development. It stems from an attachment to "the suburban lifestyle," the nearly universal preference for the detached, single-family house located a substantial distance from the site of employment. The problem is evidenced by the absence of convenient alternatives to solo automobile use. It is compounded by the lack of intergovernmental coordination of transportation and land use policies at the regional level.

WHY THE PROBLEM PERSISTS

Local governments fail to solve transportation and other critical problems because

- Participation in the problem-solving process is not broad enough. Too many stakeholders—those with the ability to make or veto decisions—do not actively and directly involve themselves in the problem-solving process. Leaders from the private sector, user groups, and contiguous governmental jurisdictions must be part of the process. There is a need for increased public- and private-sector partnership, greater user involvement, and more effective intergovernmental coordination if these types of problems are to be solved.
- No mechanism exists for consensus building. There is no forum for communication and coordination among the many participants in the process. Yet broad support is critical if any solution is to work. Regional government or the creation of other new, formal institutions is not perceived to be the answer.
- Real problem solving is not the explicit goal. Frequently, there is no problem orientation, no shared definition of the problem. Too often participants are content to merely issue lengthy planning documents instead of taking concrete steps to actually solve problems.

Ideologies get in the way, as do communication patterns and leadership styles. As a result, problems persist. The usual linear approach to problem solving is to sequentially separate planning from the processes of building support and implementing solutions. The major difficulty with this approach is that those who are left out of the process or who lose out can block or delay the implementation of any plan of action.

It is no secret that this conventional linear approach is not working. In transportation, toxic waste management, education, and a variety of other public policy areas, the more planning is done, the worse things appear to get.

Don Michael (*1*), an influential critic of traditional planning and governance models, describes the problem as one of reconciling current trends toward decentralization, autonomy, deregulation, and community-based self-help with the recognition that problems are really systemic in nature.

According to Michael, public policy problems increasingly cannot be resolved within the usual frame of reference. To what single jurisdiction can responsibility for solving transportation problems be assigned? Obviously, what one jurisdiction does affects all of the other jurisdictions in the region. Allocating resources is increasingly complex because of the recognition that one generation's allocation affects the resources available to succeeding generations. Causes and effects of problems cannot easily be isolated. There is, in other words, an urgent need to review the problem holistically or systemically.

The best way to do this is to give up the modern world view in which everything is seen as separate. In its place, Michael believes that an appropriate systemic view needs to be adopted:

> Organizations and individuals must see themselves more as 'part of' rather than 'separate from,' especially with regard to their boundaries and their task focus—in other words, their 'turf.' Autonomy should be regarded as variable, changing according to the task at hand, and people and organizations should regard themselves as able to accept or relinquish leadership. Pliability is more the precondition for survival than boundary and functional rigidity; the ability to collaborate leads to more control of outcomes than does the ability to dominate.

The ability to control outcomes is, indeed, quite limited, but present norms define competence in terms of the ability to control outcomes. Most planning efforts "preserve the illusion of control by hiding uncertainty." If uncertainty is acknowledged, however, there is an opportunity to redefine the meaning of competence. The competent leader is less someone who knows what to do. Instead he or she is a learner who enables others to learn. This new norm, Michael (*1*) believes, "transforms long-range planning from an engineering activity into a process for learning our way into the future(s)." In politics, we need to

> lessen our obsession with the adversarial, either/or, win/lose norm . . . and acknowledge that a situation is both/and, that both choices could be right or both wrong; in such cases, the sensible approach is to preserve alternatives to be tried without prejudice at a later state—without the wasted time and effort of trying to hide 'failure' or appearing to be 'right,' or the usual wasteful and demoralizing buck-passing games. When one is operating from a learning stance, one must be able to let go of or transform a commitment if it is not accomplishing what inspired it in the first place.

Michael is also critical of dependence on technical and quantitative data to define problems and possible solutions. The unquenchable thirst for "more and better data" gives rise to the illusion that the data are objective, usable, and uncertainty reducing. Indeed, they are none of these things. In a systemic world, all information is partial. It is also value loaded. Most important, no validated predictive models of social change have ever been developed. For Michael, the proper roles for the expert or technician are those of learner, educator, and process manager.

This critique is relevant to transportation problem solving for a number of reasons:

- Congestion and other transportation problems are best viewed as systemic in nature. The transportation system does not exist in a vacuum; beyond its boundaries lies a broader economic, social, political, and psychological context.
- The system for addressing transportation problems is also highly decentralized with many autonomous actors. Not only are different levels of government involved, but numerous contiguous jurisdictions throughout any metropolitan region are also affected. In addition, numerous private-sector and user interests hold a stake in the process.
- In such an environment, collaboration and consensus building have many advantages over majority vote or other adversarial decision rules.
- There is considerable uncertainty about policy outcomes and much learning to be accomplished. In transportation, the political task is, as Michael says, "to determine value priorities and to revise them as learning makes the consequences of the set of premises more clear." When there is clarity about preferred values, technical data can be used to select courses of action consistent with those values.

COLLABORATIVE PROBLEM-SOLVING MODEL

Collaborative problem solving confronts the challenge of viewing problems systemically while honoring the values of decentralization, separation of powers, limited government, and private property rights. Interest in these techniques is growing. They are being applied within organizations that need to increase productivity and manage rapid change. A number of localities have embarked upon strategic planning projects that are consensus based and collaborative in nature.

Richard Bradley (unpublished material) has identified the following general principles that are common to most collaborative problem-solving approaches:

- They are consensus based. Because any party has the power to block a decision, these projects avoid majority voting in favor of consensus-based decision making.
- All of the stakeholders must be involved. If all of the parties with the power to influence the outcome are part of the process, it is more likely that implementation will proceed quickly. In transportation, these parties include local elected officials, public-sector managers, executives with major-employer organizations, developers and other actors in the land development process, and leaders of user and citizen groups.
- They are problem oriented. Most of these approaches to problem solving involve working with the parties to reach agreement on the definition of the problem before moving on to discuss solutions. "In general, if parties don't agree on the problem, they won't agree on a solution."
- These approaches are interest based as opposed to position based. Typically, leaders try to keep parties from becoming locked into hard positions by getting them to articulate general interests and concerns before advancing proposals. "If parties can legitimize each other's rights and interests, it is much easier to find win/win solutions."

• Involvement of neutral third parties. "Because it is difficult to be a stakeholder/participant and a disinterested convener/facilitator at the same time, most processes can be made more effective by the involvement of a neutral third party known as a mediator or facilitator. Neutral auspices help to guarantee that all parties' interests will be heard and that no one is trying to manipulate the process."

Experience in Santa Clara County suggests the following additional general principles:

• In addition to neutral facilitation, there is a need for expertise in overall process design. Who meets on whose turf to discuss what agenda must be carefully planned. The way in which the project is launched, positioned, and funded is critical to securing broad-scale participation. The process must be designed by those knowledgeable about the political environment and the substantive issues being dealt with.

• Collaborative planning projects are dynamic in nature and, hence, require active management. Careful monitoring and frequent reevaluation are needed to ensure the integrity of the process. Beginning with recruitment of participants and extending throughout the process, project leaders must inspire a high level of commitment to the principles of collaborative planning.

• The principle of peer participation is important. The process works best if all participants perceive themselves to be working in groups with their peers.

• Participants must assist in designing the process in order for it to work. If the process is dominated by staff or consultants, it is likely to break down.

• Participants must agree to respect the process and not advance solutions of their own before the group's work is completed.

• Meetings must be carefully planned to build and maintain momentum. Participants need to be consulted on agendas. Issues need to be properly focused and sequenced for group attention.

• Highly skilled facilitation is required. Facilitators must be highly skilled in collaborative problem solving, committed to the process, and have no stake in the content of the solution.

• Careful management of technical information is important. Participants are typically sophisticated consumers of decision-making information. Staff must produce planning data and analyses that are at an appropriate level of detail, understandable to nontechnicians, and concise. The material should be pretested with a small sample of key participants to be sure that it is suitable.

• Although prime participants are usually a select group of opinion leaders, some program for involving the rank-and-file public will strengthen the process. Traditional "public hearings in the auditorium" formats do not provide a meaningful opportunity for participation and are best avoided.

• If real problem solving is to be accomplished, participants need to take the time to define a common mission, clarify what concrete results they want to produce, and set standards by which to assess performance. They also need to be willing to identify barriers to accomplishment as well as strategies for overcoming the barriers. When these issues are dealt with, participants can function as an effective team.

• Collaborative problem-solving projects work best if media personnel can be encouraged to work in a new way. On-the-spot, 20-sec coverage tends to contribute to a polarization of positions. However, in-depth reporting and analysis of issues can contribute to the consensus-building effort by raising public awareness of and support for constructive problem-solving actions.

The advantages of collaborative techniques are many, including faster implementation, better solutions due to the pooling of ideas, and increased enthusiasm for future problem-solving enterprises.

CASE STUDY: TRANSPORTATION 2000, SANTA CLARA COUNTY, CALIFORNIA, 1984–1987

Description of Project

The Transportation 2000 project, completed in early 1987, is a midrange transportation planning project in Santa Clara County, California—the Silicon Valley area south of San Francisco. The area is one of the most rapidly growing in the United States and has a worsening congestion problem. Transportation has emerged as the most salient local issue. The policy environment is characterized by 15 separate municipal jurisdictions within the county with a history of competition for the fiscal benefits of development. Only recently have attempts been successful in establishing a mechanism for meaningful coordination of transportation and land use policies, and this effort is in an early stage.

The purpose of the project has been to reach consensus on a plan for dealing with the mobility problem among top-level decision makers from the public as well as the private sector. Work has progressed through two phases: The objective of the first phase (1984) was to identify priority corridors for rail improvements before a deadline of the regional planning agency, the Metropolitan Transportation Commission (MTC). The objective of the second phase (1985–1986) has been to fill out the rest of the plan (i.e., develop a comprehensive transportation plan integrating rail, roadway, bus, and transportation demand and supply management elements).

The formal decision structure for the effort consists of three groups:

• Policy committee. This group directs the project and is composed of nine members: two county supervisors, three San Jose City Council members, three members of other city councils (one from each of three different cities), and one member from the general public. An additional three positions on the committee are ex officio and are held by one person from the MTC, one from the Association of Bay Area Governments, and one from the California Department of Transportation (Caltrans). Policy committee members are accountable to the five-member County Board of Supervisors.

• Citizens' advisory committee. Named before the collaborative planning process was launched, this committee consists of approximately 20 members: representatives of transportation issue advocacy groups (bicycle enthusiasts, rail transit adherents, highway advocates); business and labor organizations; good government groups (e.g., the League of Women Voters);

and transit-dependent groups (elderly, racial minorities, and handicapped).

• Technical advisory committee (TAC). This committee is composed of public works and planning staff members from the 15 cities and county governments.

Description of the Collaborative Process

• Identification of stakeholders. In addition to five elected county supervisors and policy, citizens' advisory, and technical advisory committee members, approximately 12 individuals, representing local developers and major employers, were added to the process. The rationale for the expansion was that developer and major-employer actions would be critical to solving the problem. Participants totaled approximately 70.

• One-on-one interviews. Except for TAC members, collaborative planning consultants interviewed all 70 prospective participants to elicit views on the problem, perceived causes, barriers to problem solving, and possible strategies for overcoming the barriers. The interviews were designed to elicit information, but they also began to create a context for participation in the undertaking. Participants were told that the intention was to involve top-level decision makers in an effort to solve the problem not merely discuss and analyze it. Similar one-on-one sessions were also built into the process at later stages of the project.

• Facilitated small-group work sessions. All 70 participants were divided into three small working groups of approximately 25 each. Participants were assigned to small groups randomly, except that an effort was made to ensure that each group had roughly equal numbers of private-sector leaders, government representatives, and citizen activists. A different consulting team facilitator was assigned to each of the small groups. In addition to facilitating the sessions, the facilitator's role was to serve as a liaison to nonattendees so that, as work progressed, everyone believed that he or she was part of the process. A professional recorder and technical resource person were also assigned to each small working group.

• Public involvement. Periodically throughout the process, public work sessions were held. More than 1,000 county opinion leaders were carefully identified and personally invited to attend these sessions. The list was a broad cross section of neighborhood, minority, church, and civic leaders, most of whom had never before participated in transportation planning work.

Small groups were incorporated into the design of the public work sessions. Participants spent approximately 1 hr in a large group at the beginning of the session to gain an overview of background material relevant to the items on the evening's agenda. They were broken into small groups (of 10 to 15 members) for 90-min discussions of the substantive issues. Each public workshop discussion group had a facilitator and recorder assigned to it for the evening. In the early phase of the project, the facilitators were volunteers without formal training or experience. In the later phase, only professional facilitators were used.

At the end of the discussion group sessions, participants reassembled in a large group for a brief closing session to complete the evening. The major purpose of the closing session was to give the entire body of participants consistent feedback on the points of view and opinions expressed during the evening. Sessions began at 5:30 p.m. and included a free light supper.

Between public workshops, participants received periodic reports and a regularly published newsletter.

• Information management. Great care was taken to develop appropriate technical analyses for process participants. Planning staff worked closely with consultants to frame public policy questions, key trade-offs, and relevant background data. Every effort was made to develop concise, easy-to-digest information at an appropriate level of detail. The material was mailed in advance to all persons who indicated that they were planning to attend a session.

• Use of quantitative public opinion survey. At the beginning and conclusion of the project, probability sample surveys of county citizens were completed in order to provide further guidance to participants. The surveys focused on public attitudes toward current policy initiatives, agency performance, evaluation of proposed new facilities, financing, and land use planning issues. Survey results provided useful data for comparison with the "portrait" of opinion gained at the public workshops.

• Media relations strategy. An effort was made to use the general circulation daily newspaper as a forum for community discussion about the traffic congestion issue. Project leaders met periodically with newspaper editorial staff to inform them about the project and substantive issues. As a result, newspaper staff developed considerable sophistication about transportation issues, which contributed to extensive and balanced reporting and analysis.

• Implementation strategy. In order to guarantee that concrete steps would be taken to implement the plan, participants were asked to define an implementation strategy. The strategy consisted of forwarding the plan to the cities for formal consideration and creation of an ongoing monitoring and oversight group.

Results to Date

The Transportation 2000 project was an experiment incorporating innovative approaches to collaborative problem solving. The process itself was a learning experience, and significant results have been achieved, both in transportation planning and in the development of the collaborative planning model. The following are some of the most significant accomplishments of the project:

• Development and adoption of a comprehensive long-range transportation plan for Santa Clara County;

• Expanded awareness of the institutional and political barriers to effective problem solving in transportation (e.g., absence of a regional mechanism for policy coordination);

• Appreciation of the need for an integrated, multimodal transportation system that offers choice and maximizes convenience in order to attract users;

• Consensus among leaders on new rail, bus, and roadway facilities needed in the medium term (through the years 2000–2010);

• An increased commitment to take steps necessary to solve the problem, including a willingness to levy additional taxes;

• Extensive public involvement in the planning process (more than 1,300 people attended 35 separate public work sessions);

• Increased media sophistication, coverage, and analysis of transportation and related issues; and

• Openness to future collaborative problem-solving processes including the recently established Golden Triangle Task Force, which is seeking to devise collaborative transportation demand management programs involving a number of municipalities.

CASE STUDY: FREMONT–SOUTH SAN FRANCISCO BAY ALTERNATIVES ANALYSIS AND ENVIRONMENTAL REVIEW

Description of Project

The Fremont–South Bay Alternatives Analysis, funded by the Urban Mass Transportation Administration (UMTA) and led by the MTC, the Bay Area's transportation planning agency, is an attempt to determine the most viable transit solution to worsening congestion in the Fremont to Santa Clara County transportation corridor in the South San Francisco Bay area. Work trips are projected to double in this corridor from 78,000 per day in 1980 to 144,400 per day by the year 2000. Although planned roadway improvements are expected to provide some relief, planners have concluded that it is not feasible to construct enough additional freeway lanes to meet future demand. Some type of transit solution is needed.

Nine alternative solutions have been identified. Six involve rail facilities, including possible extension of Bay Area Rapid Transit (BART) to downtown San Jose at an estimated cost of approximately $1.2 billion. The others are an improved bus system, a transit supply management approach, and a "no action" alternative.

The corridor spans two counties and six high-impact cities. When the project began in the fall of 1986, there was no consensus on a preferred transit solution. Indeed, there was considerable disagreement about how to proceed. Elected officials from San Jose, the dominant city in Santa Clara County, had stated their interest in bringing BART to their downtown. The smaller communities of Mountain View and Sunnyvale, located northwest of San Jose, favored a light rail transit system that would link their major employment sites with residential areas to the east. These smaller communities feared that their needs would be ignored at the expense of San Jose's interest in an expensive BART extension.

Given dwindling federal transit subsidies, area elected officials recognized that local consensus would be necessary if they were to be successful in winning federal support. They therefore decided to use the UMTA-mandated alternatives-analysis process as a mechanism for forging local consensus on a preferred transit strategy.

The formal decision-making body for the project is a policy committee comprised of representatives of MTC, BART, the Santa Clara County Transit District, and Caltrans. The policy committee is advised by a technical advisory committee comprised of analysts from the participating transit operators, the six high-impact cities, the Federal Highway Administration, UMTA, and Alameda/Contra Costa Transit.

Description of Collaborative Process

• Initial identification of stakeholders. In addition to elected county supervisors from each of the two affected counties, city council members from six high-impact cities were recruited to participate in the process. The cities are Fremont, in Alameda County; and Milpitas, San Jose, Santa Clara, Sunnyvale, and Mountain View, in Santa Clara County. A number of major-employer and industry organizations were also included. Citizen activists were included at public work sessions.

• Involving stakeholders in final process design. Based on the experience with Transportation 2000, an effort was made to more actively involve stakeholders in the design of the problem-solving process. One-on-one meetings were held with selected members of the policy committee and other stakeholders. As a result of these meetings, major revisions were made in the preliminary design proposed by collaborative-planning consultants. First, city managers were added to the list of stakeholders because of their key role in shaping the views of council members from their cities. Second, private-sector involvement was expanded to include all major employers as well as key "movers and shakers" within the business community.

Participants indicated in these early design conferences that the process would be strengthened if city managers could meet before the intercity council member work sessions and have a role in shaping agendas. Also, given the importance of peer participation, it was thought that business executives should meet separately from elected officials, at least in the initial stages of work. Ultimately, six city managers and the Santa Clara County Executive were identified as stakeholders, plus approximately 60 elected officials and 50 leaders from the private sector. Policy and technical advisory committee members also participated in work sessions during the course of the project.

• Facilitated small-group work sessions. The problem-solving process proceeded along three tracks: elected officials meeting as a group, private-sector leaders meeting separately as a group, and the city managers meeting on their own. City manager meetings have focused on framing agendas for work sessions of elected officials and on critiquing background analyses drafted by staff. At work sessions of elected officials and private-sector leaders, participants are divided into small working groups of approximately 12 people each. A consulting team facilitator is assigned to each of the small groups as are a professional recorder and a technical resource person. At the end of each work session, all participants reconvene in a large group for a brief report on the outcomes of the small-group work.

• Organizing the agenda. Given the substantive issues being dealt with, the process proceeded in two rounds. The first round of work sessions focused on determining what rail modes should be considered in each portion of the corridor. During this round, participants also grappled preliminarily with the staging issue: what parts of the corridor would be slated for rail service immediately and what parts would have to wait? To keep the discussion manageable, questions were posed in terms of what is best for each of three subcorridors into which the overall corridor has been divided.

The second round of sessions focused on alignment and station location issues. Using the results of these two rounds, the policy committee is to reduce the range of rail alternatives to be studied in detail. Transportation supply management and bus alternatives are automatically studied in detail, given federal rules.

• Public involvement. During each round, public work sessions were held. As was the case with the Transportation 2000 project, hundreds of opinion leaders were invited to participate. The public work sessions were configured much like those of the Transportation 2000 project.

• Information management. As was the case with the Transportation 2000 project, great care was taken to develop appropriate technical analyses for participants. The material was pretested with city managers and others and mailed in advance to participants.

• Press strategy. Consultants worked with the lead agencies on the project to develop a press strategy that encouraged in-depth reporting and an analytical approach—as opposed to focusing on the political personalities and political trade-offs of the policy process.

Results

The project has yielded the following significant results:

• The project has provided a rational process for dealing with a complex set of issues that local elected officials and citizens feel very strongly about. It has, for example, been useful to deal first with the issue of mode choice and then with alignments and station locations. The scope of discussion has been made manageable by focusing on subcorridors rather than on the corridor as a whole. City managers have played an appropriate role in the process. Affected cities have had a mechanism for dealing with a regional problem on a regional basis.

• The process has created a context of constructive communication, compromise, and accommodation that has affected the behavior of all stakeholders. Participants have attended meetings, articulated their interests, listened to the concerns of others, and attempted to deal with the problem in a cooperative way. No one has taken unilateral actions based on a narrow definition of self-interest. There has been little "posturing for negotiation's sake." One city council was recently criticized by a newspaper for having forthrightly stated a real preference for light rail transit over BART on the grounds that the council should have "held out for BART in order to strengthen their bargaining position later."

• Participants have discovered that their interests are not as divergent as they appeared to be when the project began and when perceptions of interests were based on rumor as opposed to face-to-face exchange of opinion.

• Participants have seen that there is a shared vision of regional transportation, and differences of opinion are much more about staging than about modes and alignments.

• The project has led to increased community support for transportation facility development because of extensive citizen participation in the process.

Lessons Learned

• Actors who are critical to problem solving must participate extensively in the consensus-building process, particularly top-level public-sector decision makers from affected jurisdictions. City manager participation is critical.

• Participants need to focus on fundamental issues: what is the mission and how committed are participants to problem solving?

• Participants need to move beyond planning to a genuine commitment to implementation.

• Participants need to have a sense of ownership of the process and a timeline for action. Staff and consultants must act as facilitators. Stakeholders must take ultimate responsibility for project outcomes.

• Top-level leadership is needed. Midlevel staff cannot, by themselves, lead projects of this type. A top-level executive or elected official who shares in the project vision can more easily open doors, recruit participants, and deal with the inevitable communication problems that occur in undertakings of this type.

• In order to maintain momentum, considerable time must be devoted to one-on-one work with participants who may miss meetings. Participants who want to drop out must be encouraged to continue, and facilitators must assist them in resolving their barriers to participation.

• Formal project structure can easily get in the way. The Transportation 2000 Citizen Advisory Committee duplicated many other mechanisms for citizen participation in the process and this overlap may have been confusing to some participants.

• Increased stakeholder involvement in the design of the process leads to increased legitimacy of the process and forecloses nonparticipation as an option for reluctant stakeholders. If participants participate in project design, their objections are overcome and their "buy-in" is dramatically increased.

• More effective communication takes place when groups are effectively constituted as peers. Elected officials are often reluctant to speak candidly with staff present and vice versa. Top-level managers in both the public and the private sector are often reluctant to meet with groups or individuals whom they perceive as having less organizational "clout."

REFERENCE

1. D. N. Michael. Neither Hierarchy nor Anarchy: Notes on Norms for Governance in a Systemic World. In *Rethinking Liberalism* (W. T. Anderson, ed.), Avon Books, New York, 1983, pp. 251–268.

Publication of this paper sponsored by Committee on Citizen Participation in Transportation Planning.

Management System for Repair, Evaluation, Maintenance, and Rehabilitation of Inland Water Transportation Facilities

MICHAEL J. MARKOW, DHARMA R. ACHARYA, SUE MCNEIL, AND ANTHONY KAO

Development of a management system for the maintenance and rehabilitation of locks on the inland waterways is described; the concepts and procedures apply to other facilities as well. The design of the management system is based on a life-cycle analysis of the performance and costs of facilities as affected by Repair, Evaluation, Maintenance, and Rehabilitation (REMR) policy. Life-cycle analyses of facilities require a new approach to looking at the performance of a facility and the factors that influence costs throughout its service life. This approach is referred to as "demand responsive" because maintenance, repair, rehabilitation, and reconstruction are viewed as responses to the demand for repair or renewal of a facility. Treating demand-responsive activities requires that the estimates of future resource requirements for and costs of maintaining facilities not simply be extrapolated from past trends; they must instead be based on predictions of structural and operational deficiencies caused by use, environment, and age. The concepts involved in applying life-cycle costing to analyses of REMR policy are discussed. Example models of facility performance for lock gates, walls, and mechanical equipment are developed; this performance is related to the costs and the impacts of different REMR policies; and building these models within a prototype version of a PC-based management system is discussed. The prototype REMR management system is then applied in several examples to demonstrate the application of demand-responsive maintenance concepts to realistic problems, to illustrate management system features and procedures, and to interpret system results.

The work described in this paper has been sponsored by the U.S. Army's Construction Engineering Research Laboratory (CERL) as part of the Army's Repair, Evaluation, Maintenance, and Rehabilitation (REMR) program. The REMR program responds to a growing need to keep civil works under the jurisdiction of the U.S. Army Corps of Engineers in safe, working condition. The rationale for the REMR program is supported by statistics indicating the growing importance of maintenance, repair, and rehabilitation compared with new construction. For example, operations and maintenance (O&M) have consumed rapidly increasing shares of total Corps appropriations for civil works in recent years, as the data in Table 1 (*1*) indicate.

M. J. Markow and D. R. Acharya, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139. S. McNeil, Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa. 15213. A. Kao, Department of the Army, Construction Engineering Research Laboratory, Champaign, Ill. 61820.

TABLE 1 PERCENTAGES OF CIVIL WORKS APPROPRIATIONS DEVOTED TO OPERATIONS AND MAINTENANCE VERSUS NEW CONSTRUCTION (*1*)

Year	O&M (%)	Construction (%)
1967	16	79
1970	24	66
1977	28	65
1980	35	56
1983 (estimated)	40	46
1985 (projected)	50	

The objective of this research is to develop a management system for the evaluation, maintenance, repair, and rehabilitation of civil works under the jurisdiction of the Corps of Engineers. The management system is built on the following premises:

1. REMR activities need to be seen in their economic, as well as technical, dimensions. Consequences of REMR policy alternatives (to the Corps as well as to industry) must be reduced to an economic basis for comparison.
2. Trade-offs among evaluation, maintenance, repair, and rehabilitation over time need to be accounted for.
3. Trade-offs in distributing or allocating resources among competing needs throughout a network of facilities also need to be considered.
4. In addition to the need for economic relationships in Item 1, there are questions of distribution (e.g., to whom do the costs and benefits of the REMR program accrue?) and of the influence of noneconomic decision criteria (e.g., defense needs) on the selection of the most appropriate REMR alternatives.

To scale the Corps' diverse and extensive system of projects to more manageable proportions, the current research is limited to one class of inland waterway structures: navigation locks. This approach allows exploration and development of the engineering, economic, technological, and management principles and relationships needed to address facility maintenance and rehabilitation for this type of structure. (It should be noted that locks throughout the country have significantly different dimensions, capacities, and structural and operational features.) When the applicable concepts, principles, analytic methods, and computer software have been developed and demonstrated for locks, they can be extended and adapted to other civil works within the Corps' inventory.

LIFE-CYCLE COSTING

The Corps of Engineers has had long experience in applying economic principles to engineering decisions. Calculations of benefits versus costs have been routinely applied to the evaluation of water projects for many years. These procedures extend projections of project costs and benefits through an analysis period and, by comparing the discounted totals of various alternatives, identify the economically most efficient project option or decision. As applied to the analysis of REMR projects, life-cycle costing of existing facilities considers the total costs of evaluation, maintenance, repair, rehabilitation, reconstruction, operation, use, and (in special cases) abandonment of a facility through its service life.

Demand-Responsive Approach

The implementation of life-cycle analyses of facilities required a new approach to looking at the performance of a facility and the factors that influence costs throughout its service life. This approach is called "demand responsive" because maintenance, rehabilitation, and reconstruction are viewed as responses to the demand for repair or renewal of the facility. This demand for work arises through both a physical dimension (the condition of the facility, reflecting the quality of initial design and construction; the accumulation of wear and damage from the combined effects of traffic loads, environment, and age; and corrections due to past repairs) and a policy dimension (standards of initial design and construction and the level of maintenance, repair, rehabilitation, or reconstruction to be performed, expressed through quality standards). Furthermore, because the prediction of facility condition is central to the demand-responsive approach, the impacts, as well as the costs, of alternative investment policies can be computed.

Treating REMR actions as demand-responsive activities requires that two additional elements be introduced within existing planning and management models. The first is that estimates of future resource requirements and costs cannot be extrapolated solely from past trends because these data reflect past policies and practices. Instead, the estimates must be based on predictions of structural and operational deficiencies caused by use, environment, and age as affected by future REMR policy. The second is that new relationships must be identified between the as-maintained state of the civil facility and the impacts on both the Corps and industry (the users of the facility) to provide a measure of the benefits (or disbenefits) of each policy at the costs computed. Organization of these ideas within a unified structure is shown in Figure 1.

Applications

The analytical procedures needed to implement the management structure in Figure 1 have been organized within simulation models and closed-form optimization procedures, both of which have been used to address different types of investment decisions in the transportation field. The development of simulation models is described elsewhere (*2, 3*), and the mathematical optimization procedures are presented by Fernandez-Larranaga (*4*). These tools have been applied to a diverse set of problems encompassing optimization of investments (*5–7*), evaluation of alternative investment programs (*8, 9*), allocation of scarce resources among competing activities (*10*), predicting impacts of deferred maintenance (*11*), and financing maintenance and rehabilitation (*12, 13*). Recently the optimization approach of Fernandez-Larranaga (*4*) was refined to develop simplified models and engineering curves for use by engineers in the field (*14*). Thus the demand-responsive approach provides a powerful framework for addressing decisions in facility life-cycle management, and it can be applied to a number of problems in facility investment and maintenance.

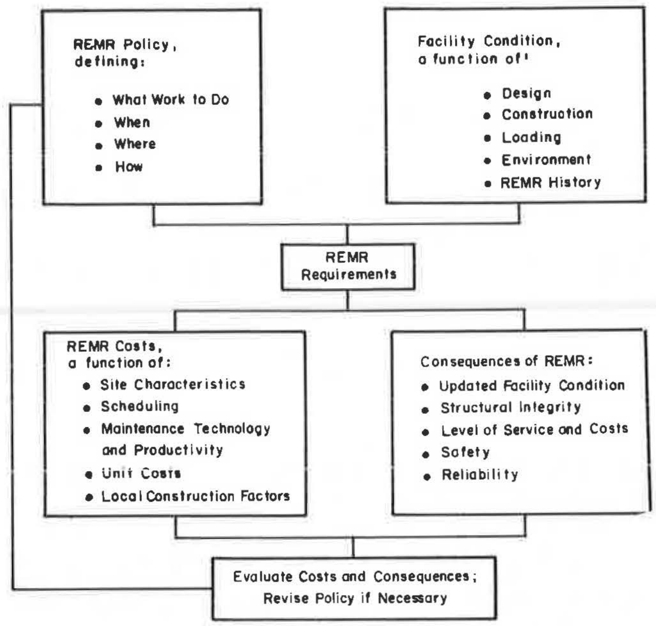


FIGURE 1 Approach to REMR planning and management.

Analyzing Life-Cycle Cost Streams

Cost streams (for both agency costs and user costs) are shown schematically for two facility strategies in Figure 2. It is assumed that traffic and environmental factors are identical in both cases but that initial facility design and subsequent performance differ in response to capital investment and maintenance policy.

These differences are evident in the respective streams of agency costs and user costs. Strategy 1 in Figure 2 entails higher agency costs for construction, maintenance, and rehabilitation but lower costs of facility usage; Strategy 2 presents the opposite pattern, with lower agency costs but higher user costs. The first strategy may be interpreted, for example, as that for a facility built and maintained to high standards to ensure premium service throughout its life. The second strategy may then be interpreted as one for a conventional facility maintained adequately but not exceptionally.

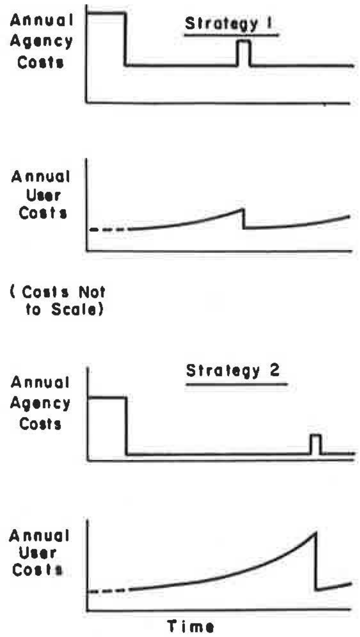


FIGURE 2 Schematic cost streams for two REMR policies.

From an agency perspective, Strategy 2 is the lower-cost alternative and perhaps would be preferred. From a total cost viewpoint, however, the savings in agency costs in moving from Strategy 1 to Strategy 2 are offset by the increase in user costs. Therefore it cannot be said a priori that one strategy is better than another; that determination depends on the relative total costs of the two options and the discount rate at which they are analyzed.

The analysis of total costs can be summarized as follows. Each of the cost streams in Figure 2 would be discounted to compute present costs. Present agency costs and present user costs would be summed in each strategy to yield net present total costs. The respective net present total costs would then be compared to identify the alternative that has the lowest total discounted costs; that alternative is the preferred option.

Evaluation of Results

The total discounted costs of a set of REMR policy options may be compared to identify the best policy, with or without budget constraints. To illustrate how this is done, assume that the benefits can be reduced to monetary terms and thus compared directly with costs. Furthermore, assume that instead of investigating only two policies, as shown in Figure 2, several policies are actually tested using a simulation model.

The results of each policy may be organized in ascending order of costs to the agency owning the facility. Because impacts or consequences of REMR policy are also in monetary terms (in this example), they can be plotted on the same graph with costs for each policy. If REMR policies are sensibly defined and efficiently carried out, more expensive policies (to the agency) should yield more advantageous impacts (i.e., greater reductions in costs associated, say, with safety, travel time, or trip reliability), which leads to the diagram in Figure 3.

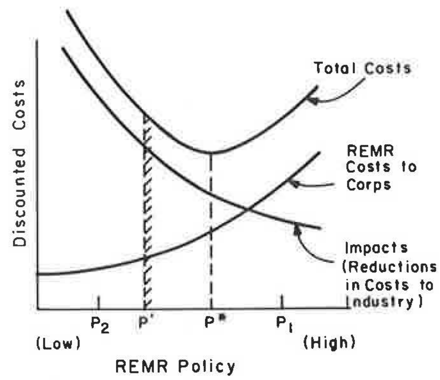


FIGURE 3 Conceptual determination of the optimal REMR policy.

Identification of the most advantageous policy now becomes a question of minimizing total transport-related costs for the network configuration and traffic specified. In the absence of budget constraints, the appropriate policy is shown in Figure 3 as P* because total costs (REMR costs to the agency plus costs associated with impacts of REMR activities) are minimized at this point. If a budget constraint is imposed, the best policy that can be funded lies to the left of P* (i.e., at P′). The relationships used to assess REMR requirements and costs, based on the principles of life-cycle costing used in the management system, are described in the following section.

PREDICTING REMR REQUIREMENTS, COSTS, AND IMPACTS FOR LOCKS

Analytic Requirements

The prediction of REMR requirements and costs according to the life-cycle framework developed in the previous section is based on the following analytic models and data:

1. Definition of measures of condition of the facility;
2. Models to predict the deterioration in condition over time, as functions of initial design and construction standards, facility age, traffic use, operating environment, and other causal factors;
3. Statements of REMR policy, expressed as quality standards defining what work is to take place, when, and where;
4. Sets of REMR activities, defining the technology to be used to correct or prevent deterioration, and the amount or quality of the improvement to be gained; and
5. Models to predict the costs and the impacts of these REMR alternatives.

Preliminary analytic models for Items 1–5 have been developed for navigation locks and are described with their results in the following subsections.

Facility Condition

Research to develop and quantify condition indices as performance measures is proceeding concurrently with this research. Therefore, how these indices must be structured to serve the objectives of a REMR management system is discussed here. It is also assumed that the measures characterize facility performance adequately and can be obtained through currently available technology for inspection and monitoring.

Furthermore, future facility condition is subject to uncertainty due to imperfect knowledge of the processes of deterioration; imperfect means of inspection, monitoring, and evaluation; and the resultant risk of unanticipated failure (such as the catastrophic failure of a supporting element). Therefore the indices used to measure facility condition in this paper have a probabilistic or stochastic dimension expressed either (a) by the mean (or expected value) of condition at some future time and the standard deviation of that estimate or (b) by the probability of failure of a lock component at some future time.

The probabilistic or stochastic aspect of facility condition is shown in Figure 4. The distribution in the upper graph in Figure 4 is based on better levels of evaluation and maintenance presumed to be applied through time T, and the distribution in the lower graph derives from a lesser policy (e.g., less frequent, or lower-quality, maintenance). For purposes of this explanation, the two cases have been constructed so that the means of the distributions at time T are the same, so that the deterioration curve follows the same path in both examples in Figure 4. The effect of the maintenance policies is then seen in the variance or standard deviations of the respective distributions, with the lesser maintenance policy presumed to result in a higher standard deviation. Essentially this means that less maintenance (or less inspection, evaluation, routine repair, etc.) of a facility leads to a loss in the reliability of its future condition.

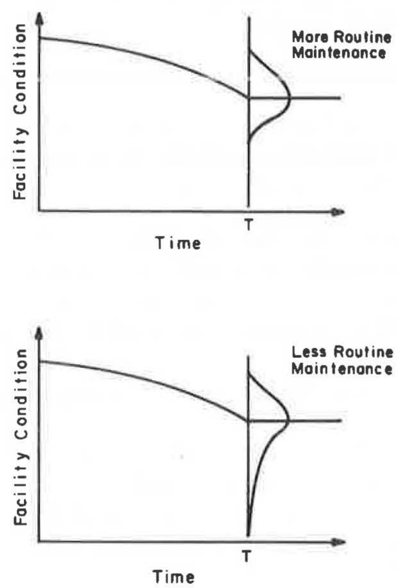


FIGURE 4 Effect of routine maintenance on reliability of facility condition.

Because locks comprise different structural and operational features, separate indices can be defined for each of these major components. There is the option of either working with this set (or vector) of condition indices or combining them (according to some empirically established formula) to compute a single overall index for the lock. In this phase of the study, the problem of expressing the condition of the lock is decomposed into separate predictions for three major components: gates, walls, and mechanical equipment.

The following measures of facility condition are used in this preliminary analysis:

- Expected value of the gate condition index,
- Standard deviation of the gate condition index,
- Expected value of the wall condition index,
- Standard deviation of the wall condition index, and
- Probability of failure of mechanical equipment.

These measures recognize the variability inherent in measures of condition (as shown in Figure 4) and the different physical and operational characteristics of several components of locks. In the examples that follow, lock gates and walls will be used to illustrate the types of preliminary relationships developed for the management system, with the understanding that analogous models have been developed for mechanical equipment. These models are intended for use only in the prototype version of the management system; the deterioration and cost models will be validated and calibrated in the field before they are included in a production version of the management system.

Deterioration Models

Three basic forms of deterioration models have been developed. The first predicts the expected values (i.e., the means) of the gate or wall condition indices. The second estimates the standard deviations of these indices over time. The third computes the probability of failure of mechanical equipment. Preliminary analytic expressions for each of these models are presented hereafter.

Expected Value of Condition Index

At this preliminary stage in the research, time has been assumed to be a surrogate for several factors that affect lock damage and deterioration: quality of design and initial construction (or of subsequent reconstruction or major rehabilitation), type and extent of lock usage, aging and time-dependent changes in material properties, and environmental effects (temperature, water intrusion, chemical attack, etc.). Subsequent research may shed light on the respective contributions of these factors to declines in the conditions of lock gates and walls and on how they can be best represented analytically. For the time being, a simple time-related function will suffice to illustrate schematically the role of deterioration functions in the REMR management system.

The expected value of the condition index for gates and walls is given by

$$CI(t) = CI_0 - a_1 \exp (b_1 \cdot t^{0.5}) \quad (1)$$

where

$CI(t)$ = condition index of the gate or wall in year t,
CI_0 = initial condition index, and
a_1 and b_1 = coefficients.

Standard Deviation of Condition Index

The standard deviation of the condition index is assumed to vary with time, the policy governing routine maintenance and evaluation, and the performance of repair and rehabilitation activities. This relationship is structured as a Markov process in which the standard deviation of the condition index in any given time period is assumed to be a function solely of the standard deviation in the preceding time period and the level of REMR activities performed in year t:

$$\sigma(t) = \sigma[t-1]\delta \quad \text{for } t \geq 1$$
$$= \sigma_0 \quad \text{for } t = 0 \qquad (2)$$

where

$\sigma(t)$ = the standard deviation of the condition index for gates or walls in year t;
δ = a variable that reflects the change in standard deviation of condition as a function of REMR activities, where $\delta > 1$; and
σ_0 = the standard deviation of the condition index in year 0.

Additional comments on the ways in which REMR activities interact to influence this relationship will be given shortly.

Effects of REMR Activities

In an analytic sense, REMR activities affect not only the values of specific variables (e.g., δ in Equation 2) but also the way in which Equations 1 and 2 must be interpreted. The reason is that activities such as repair or rehabilitation create discontinuities or steps in the deterioration functions. Thus, although the basic concepts underlying the present approach to deterioration are reflected in Equations 1 and 2, some refinements are needed to account for changes due to past REMR activities. Also, the interpretation of σ in Equation 2 needs to be more fully discussed. These extensions are covered in the following paragraphs.

Repair and Rehabilitation One major consideration in all REMR deterioration models is the effects of discontinuities in the relationship between condition and time. These discontinuities are due to activities like repair and rehabilitation, which produce an immediate and significant increase in the facility's condition index. Analytically this is important because it represents an interruption in the historical deterioration trend. The way this problem is handled in the preliminary models for locks is shown in Figure 5.

The curve in Figure 5 is a plot of the deterioration function for the mean condition index over time as given by Equation 1. The curve is interrupted by repair or rehabilitation at time T. (Whether repair or rehabilitation, and whether minor or major, would be indicated by the extent of improvement in the mean condition index.) The question is: What is the rate of deterioration after the repair or rehabilitation? The assumption made for both of these activities is that the rate of deterioration is uniquely coupled with the value of the condition index itself. This is shown graphically in Figure 5 in which the slope of the deterioration curve after repair or rehabilitation is equal to the slope of the curve at that same value of condition index (CI) before repair or rehabilitation.

This assumption is based on a concept of "equivalent facility age," which is the time between initial construction (or reconstruction or complete rehabilitation) and the time at which the condition index first intersects the reference value CI. In Figure 5, the equivalent age is denoted by m. In effect this says that the repair or rehabilitation performed at time T restores the facility to the condition it enjoyed at an earlier time m. The slope of the deterioration curve (from Equation 1) would be given by

$$-a_1 \cdot b_1 \cdot 0.5 \exp(b_1 \cdot m^{-0.5}) \qquad (3)$$

where the constants are as defined for Equation 1. Equation 1 itself can now be generalized as follows:

$$CI(t) = CI_0 - a_1 \exp(b_1 \cdot m^{0.5}) \qquad (4)$$

where m now denotes the (equivalent) age of the facility (i.e., the time since the last new construction, reconstruction, or major rehabilitation). Note that the expected value of facility

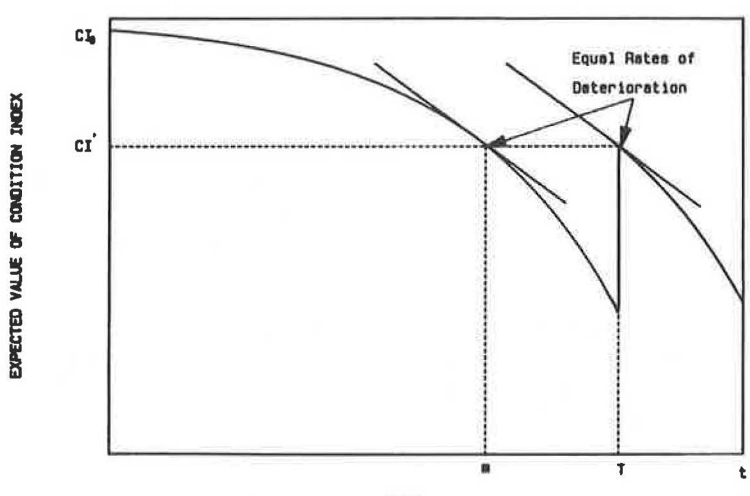


FIGURE 5 Effect of repair or rehabilitation on facility condition.

condition represented by Equation 4 captures the effects of only those REMR activities that change the magnitude of the condition index (i.e., repair and rehabilitation). It does not directly reflect the impacts of routine maintenance or evaluation. These latter activities, together with repair and rehabilitation, are reflected in the computation of the standard deviation of condition index over time in Equation 2, specifically through the variable δ.

Routine Maintenance If in some time interval there is no rehabilitation or repair, and policies governing routine maintenance and evaluation remain constant, then the variable δ likewise remains constant (and greater than 1). When routine maintenance varies, δ is a quadratic function of the difference between actual and maximum routine maintenance as given by

$$\delta = a_5 + b_5\ [\text{Max(Routine)} - \text{Routine}\ (t)]^2 \qquad (5)$$

where

a_5 and b_5 = constants,
Max(Routine) = a value representing the maximum level of routine maintenance effort, and
Routine(t) = the routine maintenance policy in year t.

Major rehabilitations restore the standard deviation of the condition index to its original value σ_0. Minor rehabilitations or repairs are assumed to reduce the standard deviation by some proportion.

Equations 2 and 5 represent, albeit in a limited but nevertheless important way, the interactions among the different REMR activities in influencing facility performance and cost. This is a characteristic of the demand-responsive approach and is important to the ability of management to assess trade-offs among different REMR policies.

Agency Costs

The cost models for routine maintenance and evaluation, lock operations, major and minor rehabilitation, and repair are based on analysis of existing data [Chapters 3 and 5 of Markow et al. (*15*) provide detail on how past cost data were used to estimate cost and deterioration model parameters]. However, two new analytic features have been introduced in accordance with concepts presented earlier: (a) REMR policies have been included specifically as variables affecting the demand for work, and hence its cost, and (b) the role of uncertainty has been explicitly recognized.

Scheduled Repair or Rehabilitation Costs

Rehabilitation costs are assumed to be proportional to an increasing linear function of the amount of improvement achieved [Section 5.2 of Markow et al. (*15*) provides justification for the linear assumption]. The relationships are as follows:

$S\_Cost(t) = 0$ if neither minor nor major rehabilitation is done in year t

$$S\_Cost(t) = \begin{cases} a_6 + b_6 \cdot \Delta & \text{if } \Delta \le \Delta_{max} \\ a_6' + b_6' \cdot \Delta & \text{if } \Delta > \Delta_{max} \end{cases} \qquad (6)$$

where

$S\_Cost(t)$ = scheduled maintenance cost for gates or walls in year t;
a_6, b_6, a_6', b_6' = coefficients;
Δ = the amount of condition index improvement in year t; and
Δ_{max} = the maximum amount of condition index improvement that can be achieved by minor rehabilitation.

Unscheduled Repair or Rehabilitation Costs

Unscheduled repairs or rehabilitations occur when the condition index is allowed to fall below a minimum standard or the mechanical equipment unexpectedly fails. For gates or walls, the expected value and standard deviation of the condition indices are known. Therefore, if the predicted value of the condition index is assumed to be normally distributed, the probability of the condition index's falling below the condition standard can be estimated and the expected cost of repair computed as follows:

$$US\_Cost(t) = inc\_prob \cdot us\_maint\_cost \qquad (7)$$

where

$US\_Cost(t)$ = the expected value of the unscheduled gate/wall repair or rehabilitation cost in year t;
inc_prob = the incremental probability of the gate/wall condition index falling below the minimum condition index standard in year t, where the increment is computed as the difference in the probability of failure in year t compared with year $t - 1$; and
us_maint_cost = the expected value of unscheduled minor rehabilitation or repair cost for gate/wall in any year.

Routine Maintenance Cost

The model of routine maintenance costs (including facility evaluation but excluding annual operating costs) is the sum of quadratic functions of the level of routine maintenance given by

$$\begin{aligned} Routine\_Maint\_Cost(t) = {} & a_8 + b_8 \cdot [G\_routine(t)]^2 \\ & + c_8 + d_8 \cdot [W\_routine(t)]^2 \\ & + e_8 + f_8 \cdot (M\_routine)^2 \end{aligned} \qquad (8)$$

where

$Routine\_Maint\_Cost(t)$ = routine maintenance cost in year t;

$a_8, b_8, c_8, d_8, e_8, f_8$ = coefficients;
$G\_routine(t)$ = routine maintenance policy for gates in year t;
$W\_routine(t)$ = routine maintenance policy for walls in year t; and
$M\_routine$ = average routine maintenance policy for mechanical equipment for the entire planning horizon.

Lock operating costs are not included here unless they can be shown to be sensitive to REMR policy.

Lock Damage Costs

Repair or rehabilitation costs due to motor vessel damage are based on the average annual damage cost because the probability of damage is assumed to be a Poisson process. A preliminary value of $15,320 per incident is used as the damage cost in the prototype management system.

Traffic Growth

Traffic volume determines both the usage of the lock (affecting REMR requirements and costs) and the impacts of lock performance as affected by REMR policy. Given the long service lives of lock facilities, it is unrealistic to expect that traffic growth can extend uniformly throughout the entire analysis period. Therefore, some growth rate may be specified for a limited number of years only; traffic will then become asymptotic to some maximum anticipated volume. The relationship used to represent traffic growth is

$$T_t = A - B \exp(-c \cdot t) \tag{9}$$

where

T_t = annual traffic in tows in year t,
A = $(1 + \text{growth rate}/100)^{\text{growth years}} \cdot$ growth start,
B = A − growth start, and
c = 0.20 (coefficient).

Growth rate is the rate of annual traffic growth in percentage; growth years is the number of years that the given growth rate is to be used; and growth start is the yearly traffic in year 0 (barges/year).

Impacts on Waterway Users

In general, REMR activities performed on locks may result in reduced shipping costs to users because of more efficient and safer performance of locks and reductions in unexpected breakdowns (although scheduled downtime may increase somewhat). Components of user (or industry) costs relevant to REMR performance include the delay cost associated with waiting in queues and lock servicing, cost incurred during facility downtime, traffic mode diversion cost, and safety and reliability cost. To illustrate these models, the delay cost model is described.

If the condition of a lock were poor, the performance of the lock would also be expected to be poor (i.e., the average service rate of the lock would be low or the standard deviation of the service time would be high, or both). As a result, traffic going through the lock would encounter longer delays. The relationship between the average service rate of the lock and the condition of the three components of the lock is captured as follows:

$$\mu(t) = \mu 0 \left| \frac{\exp\{a_3 [Mfail(t) - 0.50]\}}{1 + \exp\{a_3 [Mfail(t) - 0.50]\}} \right| \times 1/\left(1 + \exp\{b_3 [GCI(t) - 5.0]\}\right) \times \left(1 + \exp\{c_3 [WCI(t) - 5.0]\}\right) \tag{10}$$

where

$\mu(t)$ = the expected service rate of the lock in year t,
$\mu 0$ = the service rate of the lock in year 0,
a_3, b_3, and c_3 = coefficients,
$Mfail(t)$ = the mechanical device probability of failure in year t,
$GCI(t)$ = the lock gate condition index in year t, and
$WCI(t)$ = the lock wall condition index in year t.

When the average service rate and the variance in service time have been quantified, queueing theory models are used to compute the traffic delays at locks. The use of the queueing models has been common in Corps project evaluation, and a review of the existing queueing theory approaches used to compute the delays at locks appears elsewhere (*15*, Chapter 4). Whereas queueing models used by the Corps have assumed two adjacent locks to be independent of each other, new models have been developed to account for the interdependencies between adjacent locks upstream or downstream in affecting barge arrivals, delays, and departures. Thus, in a limited way, these models can account for "network" effects due to REMR work at nearby locks as well as at the lock in question. Total queueing delays computed by these models are translated into total delay costs.

Within anticipated ranges of facility condition, the effect of an REMR activity would be some improvement in the performance of an existing lock. The effect of this change in facility condition on traffic mode diversion is therefore assumed to be insignificant. Also, because there are no standard procedures for computing safety and reliability costs, such costs and benefits are not included in the management system at the present time.

Total Cost Tallies

When all of the cost items have been computed, the agency costs and user costs are totaled by summing the following items annually through the analysis horizon: the costs of both scheduled and unscheduled rehabilitation, repair, and routine maintenance of lock gates, walls, and mechanical equipment; the costs of damage to the locks by barges; and user costs due to facility downtime and to delay in the queue and during lock cycling. These annual costs are then discounted and summed for the entire analysis period.

Case Study of REMR Policies

A case study of six example REMR policies was developed to demonstrate the capabilities of the prototype management system. The six policies are given in Table 2 and range from lower-cost, lower-standard policies (beginning with Policy 1) to higher-cost, higher-standard policies (concluding with Policy 6). The higher standards reflect more frequent and more extensive REMR work performed on the facility than do the lower-standard policies. The six policies were run successively using the model; results are summarized in the cost curves in Figure 6.

The results of the case study in Figure 6 are a discrete analog to the conceptual results shown in Figure 3. For example, the least-cost solution, Policy 2, corresponds to the optimal policy P* in Figure 3. In the absence of other, noneconomic considerations, Policy 2 would therefore be the recommended REMR policy in this example. Furthermore, the trends of the individual cost curves in Figure 6 help in understanding the trade-offs among cost components leading to this result.

For example, the policies in Figure 6 are ordered such that the standards of REMR increase from left to right. Discounted agency costs likewise increase with increasing policy standards, as suggested conceptually in Figure 3. Furthermore, with more frequent and more extensive REMR work under the higher policies, average lock condition and service rate also improve, reducing the average delay per tow. The impact of this improved facility condition on user costs is somewhat more complicated, however, because a number of competing trends are at work.

User costs are relatively high for lower REMR policies because of the resulting poorer lock condition and its adverse impacts on lock cycling time and associated delays and queues. Furthermore, with the lower levels of evaluation and routine maintenance, there is a higher expected frequency of unscheduled downtime for maintenance and repairs. As the REMR policy improves, so do lock condition, service time, and delays to industry, as suggested conceptually in Figure 3. Indeed, this trend is borne out by the case study results in Figure 6, particularly if Policies 1–5 are compared. However,

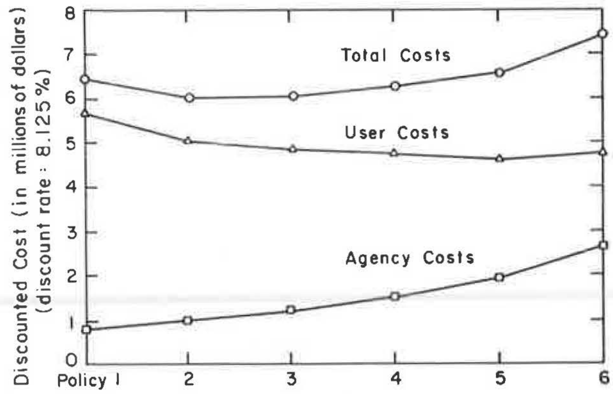


FIGURE 6 Results of case study.

although unscheduled downtime in Figure 6 decreases from Policy 1 to Policy 6, scheduled downtime increases because of the greater amount of work called for by the higher REMR policies.

At some point a limit is reached, beyond which any additional REMR work begins to interfere with lock traffic in an uneconomical way. For this case study, this trade-off is illustrated by the comparison of Policies 5 and 6 in Figure 6, at which interval the discounted user costs begin to rise, in addition to the already increasing agency costs of the higher-standard policies. The implication of these trends is that although Policy 2 is optimal, any further improvements in policy are not economically warranted: not only will incrementally higher standards cost the Corps more, they will, at some point, also delay traffic with no compensating benefit.

This comparison of competing policies (Figure 6) serves to illustrate the basic ideas and procedures involved. More important, it provides a practical example of the application of demand-responsive concepts of life-cycle costing and how these concepts can be applied in a practical way to REMR management of civil works.

CONCLUSION

The objective of this research program has been to design, develop, and illustrate a computerized package to assist districts, divisions, and the Office of the Chief of Engineers of the Corps of Engineers in the management of REMR programs for civil works. The REMR management system is based on the

TABLE 2 REMR POLICIES TESTED IN CASE STUDY

	Policy					
Item	1	2	3	4	5	6
Lock gates						
Minimum condition standard	2.1	4.1	5.4	6.7	7.6	8.8
Major rehabilitation interval (years)	49	45	40	35	30	25
Repair interval (years)	35	30	25	20	15	10
Routine maintenance level (0–10)	3	4	5	6	7	8
Repair ΔCI	0.45	0.45	0.45	0.45	0.45	0.45
Lock walls						
Minimum condition standard	7.7	8.0	8.4	8.7	8.9	9.2
Major rehabilitation interval (years)	49	45	40	35	30	25
Repair interval (years)	35	30	25	20	15	10
Routine maintenance level (0–10)	3	4	5	6	7	8
Repair ΔCI	0.2	0.2	0.2	0.2	0.2	0.2

concepts of life-cycle costing of civil facilities and demand-responsive analyses of repair, evaluation, maintenance, and rehabilitation. The research has comprised several tasks involving (a) formulation of concepts needed to manage programs of facility repair, evaluation, maintenance, and rehabilitation; (b) integration of management elements within a framework of facility life-cycle costing; (c) review of information on the frequency and costs of REMR performance and development of preliminary models of facility condition and costs; and (d) incorporation of these predictive models within a prototype REMR management system, illustration of its use, and interpretation of its results.

Results of the research to date indicate several areas in which further studies need to be made in future stages of this work. Among the more prominent topics requiring investigation are (a) a much better understanding of the mechanisms of deterioration and the role of REMR activities in correcting or preventing distress, (b) determination of appropriate condition indices at the facility level, (c) relationships quantifying the costs and impacts (or consequences) of alternative REMR policies, and (d) mathematical procedures to yield optimal REMR policies.

ACKNOWLEDGMENTS

This work has been sponsored by the Army's Construction Engineering Research Laboratory (CERL), with funding provided through the Army's program for facility Repair, Evaluation, Maintenance, and Rehabilitation (REMR). Anthony Kao of CERL has served as program monitor on this project, providing technical guidance, arranging contacts with personnel at several facilities operated by the Corps, and coordinating this work with that of other researchers in the REMR program. The authors also wish to acknowledge the contributions and suggestions of Thomas Yu of CERL.

REFERENCES

1. J. M. Scanlon, Jr., et al. *REMR Research Program Development Report.* Final Report. Office, Chief of Engineers, U.S. Army, Washington, D.C., Feb. 1983.
2. M. J. Markow and B. D. Brademeyer. *Modification of System EAROMAR: Final Technical Report.* FHWA/RD-82/086. FHWA, U.S. Department of Transportation, April 1984.
3. M. J. Markow et al. *Intercity Transportation in Egypt: Summary Technical Report.* Cairo University–MIT Technology Adaptation Program, Cambridge, Mass., June 1984.
4. J. E. Fernandez-Larranaga. *Optimal Dynamic Investment Policies for Public Facilities: The Transportation Case.* Ph.D dissertation. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, June 1979.
5. M. J. Markow. Economically Efficient Pavement Design and Rehabilitation Policies Under High Traffic Volumes. *Proc., Tenth Annual Meeting of the PTRC,* London, England, July 1982.
6. T. L. Friesz and J. E. Fernandez. A Model of Optimal Transport Maintenance with Demand-Responsiveness. *Transportation Research,* Vol. 13B, No. 4, May 1979, pp. 317–339.
7. J. E. Fernandez and T. L. Friesz. Influence of Demand-Quality Inter-relationships on Optimal Policies for Stage Construction of Transportation Facilities. *Transportation Science,* Vol. 15, No. 1, Feb. 1981, pp. 16–31.
8. M. J. Markow et al. *Analysis of Track Maintenance and Rehabilitation Policy.* MIT Center for Construction Research and Education, Cambridge, Mass., Oct. 1982.
9. M. J. Markow. Planning and Management Alternatives in Rehabilitation of a National Rail System. *Organizing and Managing Construction: Proc., 4th International Symposium on Organization and Management of Construction,* Vol. 3 (Research), International Council for Building Research Studies and Documentation, University of Waterloo, Waterloo, Ontario, Canada, April 1984.
10. M. J. Markow. *Demand-Responsive Highway Maintenance Budgeting: A Report on the Massachusetts Maintenance Management System.* Byrd, Tallamy, MacDonald and Lewis, Falls Church, Va.; the Commonwealth of Massachusetts, Dec. 1978.
11. M. J. Markow. Impacts of Deferred Pavement Maintenance. *Proc., Second Transportation Research Workshop,* Tempe, Ariz., Vol. 2, Nov. 1982.
12. M. J. Markow and T. K. F. Wong. Life-Cycle Pavement Cost Allocation. In *Transportation Research Record 900,* TRB, National Research Council, Washington, D.C., 1983, pp. 1–10.
13. U.S. Department of Transportation. *Section 137 Studies: Climate and Interstate 4R Apportionment.* Report to the United States Congress, Nov. 1983.
14. W. S. Balta and M. J. Markow. *Demand-Responsive Approach to Highway Maintenance and Rehabilitation,* Vol. 2: *Optimal Investment Policies for Maintenance and Rehabilitation of Highway Pavements.* DOT/OST/P-34/87/054. Office of University Research, U.S. Department of Transportation, June 1985.
15. M. J. Markow, S. McNeil, D. Acharya, and M. Brown. *Management System for Repair, Evaluation, Maintenance, and Rehabilitation of Civil Works.* Report 86-8. Center for Construction Research and Education, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge; U. S. Army Construction Engineering Research Laboratory, Champaign, Ill., June 1987.

Publication of this paper sponsored by Committee on Inland Water Transportation.

Prospects for Container-on-Barge Service on the Mississippi River

JAMES CREW, ANATOLY HOCHSTEIN, AND KEVIN HORN

Container-on-barge service represents an intermodal transport operation that takes advantage of high-capacity, low-cost inland waterways for the shipment of containers to coastal ports for transfer to ocean-going vessels. The feasibility of container-on-barge service between inland cities in the Midwest and the Port of New Orleans via the Mississippi River system is examined. It is concluded that, because of the significantly longer transit time for containers shipped by barge, relative to rail service, container-on-barge service will be unable to compete for time-sensitive cargoes. To succeed, the container-on-barge service will need to attract neobulk and relatively low-value containerized shipments and reposition empty containers.

This paper is based on the results of research conducted by the Louisiana State University (LSU) Ports and Waterways Institute for the Office of University Research, Maritime Administration. Examined are the possible market size and scope for container-on-barge (COB) services, using the Mississippi River Valley and the Port of New Orleans as an example. The major objective of the analysis was to assist waterway operators who are considering establishing COB services by assessing factors and conditions necessary for successful COB ventures. Intermodal rail rates and COB costs were examined to determine the trade flows that could potentially support COB services on the Mississippi River–Gulf Intracoastal Waterway System. Interviews were conducted with various port and waterway industry personnel to determine their attitudes toward implementation of COB services. A brief review of intermodalism is also presented to indicate the physical distribution requirements that COB must fulfill.

DOMESTIC HINTERLANDS

An assessment of the potential market area for COB services requires analysis of routing possibilities between major inland ports and overseas ports. The geographic scope of COB service was delineated using a transportation cost analysis of inland and ocean routings between major cities adjacent to the domestic shallow-draft waterway network and world trade regions via Atlantic, Gulf, and Pacific coast ports. The Intermodal Transportation Costing Model developed by the LSU Ports and Waterways Institute was used to compute ocean transport costs between major U.S. ports and representative ports in 10 major world trade regions. The following trade regions and representative ports were selected:

Trade Region	*Representative Port*
Mexico and Central America	St. Tomas, Guatemala
Caribbean	Kingston, Jamaica
East Coast of South America	Santos, Brazil
West Coast of South America	Callao, Peru
Northern Europe	Rotterdam, The Netherlands
Southern (Mediterranean) Europe	Leghorn, Italy
Asia	Singapore, Singapore
Australia and Oceania	Wellington, New Zealand
Western Africa	Dakar, Senegal
Southern and Eastern Africa	Durban, South Africa

Ocean costs for 40-ft containers were computed between representative ports of major world trade regions and the domestic ports of New York, Baltimore, Norfolk, Charleston, Savannah, Miami, New Orleans, Houston, Long Beach, Oakland, and Seattle. Rail rates for marine containers between major U.S. cities contiguous to the Mississippi River system and ports were compiled from published intermodal circulars. Most steamship lines and freight forwarders have privately negotiated lower volume-incentive mini- and microbridge contract rates. To reflect this situation, published nonnegotiated intermodal rail rates were discounted using parameters supplied by large container vessel operators.

The LSU Intermodal Transportation Costing Model selected the lowest combination of inland rail rates and ocean costs between domestic inland cities and world trade regions. The model also computed inland/water freight cost differentials for competing ports. COB service on the Mississippi River system would likely be to New Orleans or possibly Mobile. To assess the geographic scope of the COB hinterland, the model was used to compare transportation costs from inland cities to world trade areas through the Port of New Orleans with those of routings through other U.S. ports.

Table 1 gives the land/water freight cost differentials between New Orleans and competing ports. The land/water freight cost differentials indicate the competitive position of New Orleans for marine containers between domestic cities and world trade regions. For example, containers to and from Cincinnati can be moved through competing ports at costs ranging between $50 and $176 less per box than through New Orleans. Containers between Memphis and world trade regions can be shipped through New Orleans at lower costs than through other ports, however. The competitive advantage of New Orleans ranges between $18 and $140 a container for Memphis traffic, depending on the specific trade area served.

The data in Table 1 indicate that a Mississippi River COB service would encounter significant competition from other

Ports and Waterways Institute, Louisiana State University, 60 University Lakeshore Drive, Baton Rouge, La. 70803-7513.

TABLE 1 FREIGHT COST DIFFERENTIALS BETWEEN NEW ORLEANS AND OTHER MAJOR CONTAINER PORTS ($/BOX)

	World Trade Area									
	Central America and Mexico	Caribbean	East Coast of South America	West Coast of South America	Northern Europe	Southern (Mediterranean) Europe	Asia	Australia and Oceania	Western Africa	Southeast Africa
Representative port	St. Tomas	Kingston	Santo	Callao	Rotterdam	Leghorn	Singapore	Wellington	Dakar	Durban
Domestic ports										
St. Paul	68	24	–20	48	–66	–68	–94	82	–48	2
Chicago	68	24	–20	48	–66	–68	–94	82	–48	2
Peoria	100	78	50	110	35	34	–10	98	48	77
St. Louis	110	88	60	120	44	42	–2	106	56	84
Cincinnati	–54	–82	–112	–50	–146	–128	–176	–64	–116	–88
Louisville	46	18	–10	50	–46	–48	–74	36	–28	14
Omaha	0	2	22	18	50	54	–6	38	38	88
Kansas City	30	32	52	48	74	84	44	68	68	118
Chattanooga	–54	–82	–110	–50	–124	–124	–172	–64	–114	–86
Memphis	136	108	80	140	66	66	18	126	76	104

ports in attempting to divert traffic to New Orleans. For example, unless a COB service to Cincinnati could reduce the domestic portion of container transport costs by at least $50, no traffic would be diverted to New Orleans. To attract significant volumes of containers from Cincinnati would require savings in excess of $200 per box. In markets where New Orleans has a relative cost advantage, such as Memphis, the demand for COB service will be a function of total distribution cost savings relative to existing rail and truck service.

Unless COB can offer very low rates for low-value neobulk commodities, the overall weak competitive position of the Port of New Orleans in major midwestern river cities such as St. Paul and Chicago will remain unchanged. With the exception of a few markets located close to the port, such as Memphis, New Orleans does not have a large "captive" hinterland. This situation is even more extreme for other potential COB coastal ports such as Mobile. Moreover, existing containerized marine traffic that moves between major river cities and the Port of New Orleans is quite limited. Interviews with representatives of major railroads serving New Orleans indicate that the number of marine containers moved through the port from the major Mississippi River cities is relatively small. Almost 80 percent of the marine containers between the major river cities and world trade regions that pass through New Orleans originates or terminates in Chicago, St. Louis, or Memphis. The estimated numbers of marine containers handled annually by railroads through the Port of New Orleans are as follows: St. Paul, 500; Chicago, 5,400; Peoria, 400; St. Louis, 14,200; Cincinnati, 50; Louisville, 2,000; Omaha, 200; Kansas City, 5,000; Chattanooga, 50; and Memphis, 13,700.

MARKET AGENTS

A major determinant of the success of COB in Europe and the Pacific Northwest has been the ability of transportation agents to structure COB as an intermodal service. The perceptions of steamship lines, towing companies, and port and terminal operators are summarized for each market participant.

Representatives of steamship lines are generally quite skeptical about the feasibility of COB service:

- The speed, frequency, and cost of rail service were regarded as overwhelming any potential line-haul transportation cost savings that might arise from COB. Conventional container traffic is not regarded as divertible to slow and infrequent COB service.
- Low-value neobulk cargo volumes are perceived to be insufficient to justify regular COB service. Infrequent flows can be unprofitable for steamship lines because of expenses associated with maintaining a chassis pool at interior ports.
- COB was not viewed as a viable alternative to rail unless large steady flows of non-time-sensitive cargo could be containerized.

Towing companies are enthusiastic about establishing COB services:

- Operators believe that COB can be conveniently accommodated with existing equipment and within existing operational practices of towing companies.
- Towing companies are reluctant to accept any responsibility for cargo damage or for any non-line-haul cost components associated with COB, an attitude that goes against the trend toward intermodal pricing.
- Operators expect to be reimbursed for barge line-haul costs and all associated expenses such as tow makeup and breakup, regardless of the number of boxes available.
- The time-sensitive nature of most containerized cargoes and the importance of non-line-haul logistics costs, particularly terminal expenses, are not readily perceived by towing companies.
- Consequently, towing companies are the most vigorous supporters of an intermodal COB service sponsored by steamship lines or third parties.

Port and terminal operators familiar with COB are primarily concerned with loading and unloading sequences for containers:

- Adequate supply of chassis, yard space, damage control, and inspection of marine containers are important considerations. These activities result in perceptions of low productivity and high labor costs to load containers on and unload containers from barges.

- In some instances terminal operators do not have adequate equipment to efficiently handle marine containers of different sizes.
- Estimates of cost to load and unload marine containers varied widely among terminal operators. Quotations exceeded $100 a box to load or unload at coastal ports. Interior ports quoted handling rates of between $50 and $100 a container per move. These estimates do not include chassis costs.
- It appears that traditional public port and terminal operator container-handling practices cannot be used if COB service is to be economically viable.

INTERMODAL REQUIREMENTS

To fully comprehend the challenges faced in implementing a COB service, it is important to understand the nature of intermodal services and pricing. Although these challenges are not insurmountable, as evidenced by successful COB services in the Pacific Northwest and Europe, any new services must employ competitive intermodal service practices and pricing policies. Significant institutional changes have occurred in domestic and offshore transportation since 1980. Railroad transportation of trailers and containers has been completely deregulated. Interstate motor freight transportation is almost totally deregulated. The Shipping Act of 1984 allows steamship lines to quote through intermodal rates to interior points without distinguishing between domestic and water rates.

These sweeping institutional changes characterize a most competitive market in which railroads have a great deal of pricing flexibility for intermodal traffic. Steamship lines have contracted with railroads to obtain low volume-incentive rates. Steamship lines are increasingly moving containers on a single through bill-of-lading, bypassing freight forwarders and other third parties. Shippers increasingly expect to deal with one party for a complete service package, including responsibility for loss and damage and meeting delivery commitments. COB can only fit into emerging integrated domestic-foreign intermodalism if it is supported by steamship lines or is able to independently offer shippers sufficient real cost savings to entice them to forgo delivered prices and door-to-door service commitments of one party under the rail-water minibridge and microbridge rates quoted by steamship operators.

Railroads and steamship lines are exploiting economies of scale by building volume at a limited number of interior hub terminals and load-center ports. Volume-incentive intermodal contract rates negotiated by steamship lines and freight forwarders are between 10 and 30 percent less than nonnegotiated intermodal rates. Volume-incentive intermodal rates between selected inland cities and coastal ports are given in Table 2. The spread between nonnegotiated and estimated volume-incentive rail rates is significant relative to projected cost savings for COB. For example, the Leaseway Transportation Corporation's COB concept of saving shippers approximately $100 per container appears to be viable compared with nonnegotiated rail intermodal rates. The lower volume-incentive rates for large shippers, however, erase any appreciable COB line-haul rate savings.

New developments in rail intermodal equipment have resulted in lightweight articulated flatcars capable of handling two tiers of containers. Double-stack rail cars reduce line-haul costs between 25 and 40 percent depending on train size, length of haul, and route characteristics. Table 3 gives a projection of potential double-stack rail rates for the COB hinterland of major Mississippi River cities. Although double-stack service may never be instituted in some of these markets because of insufficient unit train volumes of containers, the overall thrust of double-stacking is negative for COB and ports not served by this technology. The absence of double-stack service between the Midwest and the Port of New Orleans is also indicative of the low volumes of containers, particularly 40-ft boxes, handled between Chicago and St. Louis and the Gulf. Existing volumes of container flows through the Gulf ports are inadequate to justify double-stack rail service notwithstanding COB. Steamship lines are repositioning their vessels to minimize port calls. Larger fourth-generation jumbo container ships are now calling at Atlantic and Pacific Coast ports. Together with steamship companies' double-stack rail cars, these vessels are pulling cargo away from small ports, aided by through rates and faster service. If COB is to be successfully interjected into the emerging intermodal hub and load-center operations, it must offer substantial savings to both steamship lines and shipprs. Discussions with representatives of steamship lines indicate that total COB logistics costs will need to be significantly lower than rail rates in order to induce shippers to forgo fast, frequent, dependable rail service. Moreover, unless COB can be operated as an extension of liner service, as is double-

TABLE 2 INTERMODAL RAIL VOLUME-INCENTIVE RATES ESTIMATED FOR MINI- AND MACROBRIDGE

	New Orleans	Houston	Savannah	Norfolk	Baltimore	New York	West Coast
St. Paul	630	750	800	650	730	780	1,160
Chicago	400	600	570	420	500	550	1,050
Peoria	350	440	450	480	550	600	1,125
St. Louis	350	440	450	480	550	600	1,125
Cincinnati	460		390	420	530	570	1,125
Louisville	350		380	390	390		1,090
Omaha	580	560	760				900
Kansas City	460	470	610	700			1,010
Chattanooga	290		220	360			
Memphis	250	400	370	540			1,125

NOTE: Plan III for single-container shipments between major COB river cities and domestic ports ($/box).
SOURCE: Compiled by LSU Ports and Waterways Institute.

TABLE 3 INTERMODAL RAIL VOLUME-INCENTIVE RATES ESTIMATED FOR DOUBLE-STACK MINI- AND MACROBRIDGE

	New Orleans	Houston	Savannah	Norfolk	Baltimore	New York	West Coast
St. Paul	470	560	600	490	550	585	1,000
Chicago	300	450	430	315	375	410	900
Peoria	260	500	340	360	410	450	975
St. Louis	260	500	340	360	410	450	975
Cincinnati	350		300	315	375	425	975
Louisville	280		300	300	300		975
Omaha	435	420	570				810
Kansas City	350	360	450	520			900
Chattanooga	230		180	180			
Memphis	200	300	300	390			975

NOTE: Plan III for single-container shipments between major COB river cities and domestic ports ($/box).
SOURCE: Compiled by LSU Ports and Waterways Institute.

stack rail intermodal equipment, shippers' commitments to use COB will not be obtainable for modest cost savings at the expense of single billings and centralized responsibility for delivery.

CONTAINER-ON-BARGE COSTS

Successful COB services in the Pacific Northwest and Europe have rate structures and service patterns that are competitive with other intermodal alternatives. To assess the prospects of implementing additional COB services on the U.S. inland waterways, an examination of towing costs and operating practices on the Mississippi River–Gulf Intracoastal Waterway (MR-GW) was conducted. This assessment formed the basis for estimating COB line-haul costs. Although the costs are based on operations centered at the Port of New Orleans, the results should have general applicability at least to the extent of providing basic information on relative competitive conditions and volumes necessary for a profitable COB service.

To provide a basis of comparison with existing intermodal services, COB line-haul costs were converted into per box costs, based on different levels of barge capacity utilization, and terminal costs (both inland and ocean). COB costs per box were then compared with rail intermodal rates to determine possible operational savings. A parametric analysis of other factors that would influence the actual costs of COB service was conducted. An evaluation of dedicated versus general towing was conducted to determine the volumes for which a high-speed, reliable COB service could be established. In-transit inventory carrying costs were examined to determine their effect on the break-even number of boxes that COB service would require to provide sufficient cost savings to attract shippers. Overhead costs were computed and combined with estimated COB operating costs to determine the volumes of containers necessary to sustain service and possible vessel itineraries.

COB LINE-HAUL TOWING COSTS

General towing charges for COB were estimated on the basis of quotations from operators between New Orleans and the inland ports of St. Paul, Chicago, St. Louis, Cincinnati, Memphis, and Houston. General COB towing costs for a jumbo barge (195 × 35 ft) are given in Table 4. The costs include fleeting, switching, and tow makeup and breakup charges of $900 per barge for movements to and from St. Paul and Chicago, and $600 per barge for the other inland ports. On the lower Mississippi, south of St. Louis, towing costs vary by direction. The costs in Table 4, however, reflect average one-way charges for round-trip barge movement.

Towing costs will not change as a function of the number of boxes carried by the barge. Average line-haul towing costs per container will, therefore, be determined by the number of boxes loaded on each barge. General towing costs per barge are divided by three levels of barge capacity utilization in Table 4:

1. Full capacity—72 containers,
2. Three-quarter capacity—54 containers, and
3. One-half capacity—36 containers.

Although average line-haul towing costs per container appear relatively low, barge utilization is very important in determining the cost per box. The estimated capacity of 72 containers is based on three tiers of 20-ft boxes. Each tier would accommodate 24 boxes in a 195- × 35-ft jumbo open-hopper river barge.

COMPARATIVE RAILROAD INTERMODAL RATES

Railroad intermodal rates are normally quoted on a ramp-to-ramp basis, which includes loading, unloading, and line-haul services. To compare COB costs per box with railroad intermodal rates for marine containers, it is necessary to add loading and unloading costs to COB line-haul costs. COB terminal costs for loading and unloading will be heavily influenced by labor rates, work rules, and productivity. To account for the prospective variability in terminal costs, COB line-haul costs in Table 4 were increased to incorporate three projected levels of per move container loading and unloading costs:

- Low cost—$30 at interior ports and $30 at coastal ports,
- Moderate cost—$30 at interior ports and $55 at coastal ports, and
- High cost—$55 at interior ports and $55 at coastal ports.

COB dock-to-dock costs per box were then compared with rail intermodal ramp-to-ramp rates for single-container shipments in Tables 5 and 6. Rail intermodal rates are specified for two

TABLE 4 LINE-HAUL GENERAL TOWING COSTS FOR BARGES BETWEEN INLAND PORTS AND NEW ORLEANS

	Towing Cost ($)	Average Container Cost ($/box) Full Capacity (72 boxes)	Three-Quarter Capacity (50 boxes)	One-Half Capacity (36 boxes)
St. Paul	10,000	139	185	278
Chicago	7,800	108	144	217
St. Louis	5,100	71	94	142
Cincinnati	7,100	99	131	197
Memphis	3,000	42	56	83
Houston	4,000	56	74	111

SOURCE: Computed by LSU Ports and Waterways Institute.

TABLE 5 COB DOCK-TO-DOCK LINE-HAUL COSTS VERSUS VOLUME-INCENTIVE RAIL MINI- AND MICROBRIDGE INTERMODAL RATES

	Volume-Incentive Rail Rate	Low Terminal	Savings	Moderate Terminal	Savings	High Terminal	Savings
72 Boxes per Barge ($/box)							
St. Paul	630	199	431	224	406	249	257
Chicago	400	168	232	193	207	218	182
St. Louis	350	131	219	156	194	181	169
Cincinnati	460	159	301	184	276	209	251
Memphis	250	102	148	127	123	152	98
Houston	250	116	134	141	109	166	84
54 Boxes per Barge ($/box)							
St. Paul	630	245	385	270	360	295	335
Chicago	400	204	196	229	171	254	146
St. Louis	350	154	196	179	171	204	146
Cincinnati	460	191	269	216	244	241	219
Memphis	250	116	134	141	109	166	84
Houston	250	134	116	159	91	184	66
36 Boxes per Barge ($/box)							
St. Paul	630	338	292	363	267	388	242
Chicago	400	277	123	302	98	327	73
St. Louis	350	202	148	227	123	252	98
Cincinnati	460	257	203	282	178	307	153
Memphis	250	143	107	168	82	193	57
Houston	250	171	79	196	54	221	29

SOURCE: Computed from Tables 2 and 4 and assuming low, moderate, and high terminal costs of $60, $85, and $110 per box.

levels: (a) volume-incentive minibridge and microbridge rates developed from interviews with people from steamship lines and railroads (Table 5) and (b) estimated double-stack rates if 100 platform unit trains (200 forty-foot containers per train) were feasible between the COB hinterland and New Orleans (Table 6).

COB dock-to-dock costs per box were subtracted from rail ramp-to-ramp container rates to indicate the operational cost advantage between water and rail service. Neither COB nor rail includes drayage costs. The operational cost advantage also does not include any allowance for increased inventory costs associated with slower, less frequent water service or COB overhead costs. Table 5 (volume-incentive rail rates) provides the best indication of current COB potential cost savings for non-time-sensitive freight (zero inventory holding costs). If barges can be loaded to at least one-half capacity, 36 boxes, COB offers the potential for substantial savings to shippers from distant interior points, such as Cincinnati and St. Paul, if inventory costs are negligible.

With incremental transit times approaching 10 and 20 days between these two ports and New Orleans, respectively, even low-valued commodities with inventory costs of $10 per day would largely negate any line-haul savings. Based on the comparative port analysis given in Table 1, Cincinnati and St. Paul are outside the market area for New Orleans except for southern hemisphere traffic. Only limited amounts of containerized traffic and a small amount of break-bulk traffic move between these cities and the Port of New Orleans. Although the cost analysis indicates that savings could be used to attract non-time-sensitive freight to COB for these two areas, existing volumes for these two ports are unlikely to be sufficient to support COB services.

As the number of containers per barge is increased, the COB competitive advantage expands to cities closer to the Port of

TABLE 6 COB DOCK-TO-DOCK LINE-HAUL COSTS VERSUS ESTIMATED DOUBLE-STACK RAIL INTERMODAL RATES

	Double-Stack Rail Rate	Low Terminal	Savings	Moderate Terminal	Savings	High Terminal	Savings
72 Boxes per Barge ($/box)							
St. Paul	470	199	271	224	246	249	221
Chicago	300	168	132	193	107	218	82
St. Louis	260	131	129	156	104	181	79
Cincinnati	350	159	191	184	166	209	141
Memphis	200	102	98	127	73	152	48
Houston	200	116	84	141	59	166	34
54 Boxes per Barge ($/box)							
St. Paul	470	245	225	270	200	295	175
Chicago	300	204	96	229	71	254	46
St. Louis	260	154	106	176	81	204	56
Cincinnati	350	191	159	216	134	241	109
Memphis	200	116	84	141	59	166	34
Houston	200	134	64	159	41	184	16
36 Boxes per Barge ($/box)							
St. Paul	470	338	132	363	107	388	82
Chicago	300	277	23	302	(2)	327	(27)
St. Louis	260	202	58	227	334	252	8
Cincinnati	350	257	93	282	68	307	43
Memphis	200	143	57	168	32	193	7
Houston	200	171	29	196	4	221	(21)

SOURCE: Computed from Tables 3 and 4 and assuming low, moderate, and high terminal costs of $60, $85, and $110 per box.

New Orleans. A threshold savings of approximately $100 per box is definitely feasible at between 50 and 72 containers per barge between Chicago and St. Louis and New Orleans if inventory and COB overhead costs are disregarded. Although this result is at odds with the short length-of-haul evidenced by existing COB operations, it reflects the more intense level of competition at inland points such as St. Louis and Memphis and the disregard of inventory and overhead costs.

PARAMETRIC ANALYSIS

One of the key features of intermodal service is the provision of rapid, reliable service that provides flexibility to shippers. Existing intermodal alternatives for shippers normally include 1- or 2-day transit times between coastal ports and inland cities at competitive rates. Relatively fast and frequent service also affects other physical distribution costs such as insurance and inventory carrying costs. As evidenced by existing COB services in Europe and on the Columbia-Snake system, the success of COB hinges not only on developing a competitive rate structure but also on functioning as a truly intermodal operation. In this section operational and cost parameters that would influence COB services and costs are examined. Of particular concern are the potential for establishing a regularly scheduled, dedicated COB tow service, the impact of in-transit inventory carrying costs, and overhead costs associated with establishing and maintaining COB services.

Dedicated Versus General Towing

As an alternative to general towing, the costs for dedicated tows providing rapid and reliable service were examined. Dedicated tows would allow a tightly scheduled operation. One advantage of a dedicated tow is the possibility of using a large towboat to increase tow speed and increase the number of trips per week. However, speed restrictions (obstacles and channel depth) limit the potential of dedicated tows on the lower Mississippi and Gulf Intracoastal Waterway. To provide a basis of comparison with costs of general towing, towing costs for dedicated COB tows were computed for New Orleans–Memphis and New Orleans–Houston itineraries.

Average one-way costs per barge based on weekly service with a two-barge dedicated tow were $7,300 between New Orleans and Memphis and $6,000 between New Orleans and Houston. Line-haul costs per box for dedicated tows are given in Table 7. If volume is sufficient, the higher cost of dedicated towing may be justified when multiple barges of containers can be moved in one tow (compared with the one-way general towing cost of multiple barges of containers in one tow). For example, the one-way general towing cost (Table 4) for two barges between Memphis or Houston and New Orleans would be $6,000 (2 × $3,000) and $8,000 (2 × $4,000), respectively. The line-haul cost of dedicated two-barge tows would remain greater than that of general towing on the Mississippi to Memphis. Dedicated two-barge tows could be as much as 25 percent less costly than general towing, however, on the Gulf Intracoastal Waterway between New Orleans and Houston (Table 4 versus Table 7).

Overhead Costs

Although COB appears to have some significant cost savings over conventional intermodal service, the previous analysis

TABLE 7 COB LINE-HAUL DEDICATED TOWING COSTS PER BOX FOR TWO-BARGE TOWS: ONE ROUND TRIP A WEEK BETWEEN MEMPHIS–NEW ORLEANS AND HOUSTON–NEW ORLEANS

	Memphis–New Orleans	Houston–New Orleans
Full capacity (144 boxes)	51	42
Half capacity (72 boxes)	101	83

SOURCE: Computed by LSU Ports and Waterways Institute.

excluded two important cost components, overhead and inventory costs. The overhead costs associated with conventional intermodal service are reflected in the pricing structure. To provide an accurate comparison, overhead costs must be added to COB charges.

COB overhead costs were estimated for a single interior port service assuming a manager, coastal port captain, interior port director, and two clerical personnel. COB overhead costs were estimated to be $6,000 per week. The break-even number of boxes for scheduled general COB towing service between one hinterland port and New Orleans is given in Table 8. Break-even volumes reflect low container-handling costs for volume-incentive rail mini- and microbridge rates and potential double-stack rail intermodal rates. (With moderate and high terminal costs, break-even volumes would increase between 5 and 20 percent and 11 and 55 percent, respectively.) COB break-even volumes for conventional rail intermodal rates increase from 40 percent of barge container capacity to 70 percent as the distance between hinterland ports and New Orleans decreases. If railroads initiated double-stack intermodal service or reduced existing volume-incentive rates, COB single-barge service would not be feasible for Chicago and Houston.

It should be noted, however, that, if COB could deliver directly to the marine terminal, additional cost savings vis-à-vis rail-truck delivery would be available and a higher COB rate would be possible. Also, as was previously noted, St. Paul and Cincinnati are not prime markets for New Orleans. When prospective COB rates are adjusted to reflect competition from other ports (Table 1), break-even utilization increases to 56 and 65 percent, respectively, for the two ports. With double-stack rail rates, COB break-even utilization increases to 86 and 126 percent, respectively.

The overhead costs of COB at New Orleans were shared with two inland ports (Table 9) to indicate the impact of direct service between two ports and New Orleans. The overall impact of multiple-port service on break-even volumes is rather substantial. Break-even volumes range between 32 and 50 percent of barge capacity under conventional rail service and rates. Table 10 gives the impact of direct service to three inland ports. Overall, only a small decline in break-even volumes is evidenced vis-à-vis two-port service. The data in Tables 9 and 10 indicate that COB service should be between multiple inland ports in order to spread the overhead costs associated with the ocean port connection. Moreover, the data indicate that a successful COB service will require high levels of capacity utilization.

In-Transit Inventory Carrying Costs

The last major cost element that has not been evaluated is the time value of the container and cargo. It has been generally assumed that COB must be oriented toward non-time-sensitive cargoes. However, the previous analysis assumed that the container itself had no associated time costs. Given the cost of containers and the effective lower utilization that will result from COB transport, this assumption has a tendency to bias the earlier analysis in favor of COB services.

To examine the effects of line-haul transit time on the relative competitive position of COB, per day carrying costs of $10, $20, and $30 per box were assumed. Transit time by barge was estimated on the basis of tow speeds and distances. Rail transit time to and from New Orleans was estimated to be 3 days for St. Paul; 2 days for Chicago and Cincinnati; and 1 day for St. Louis, Memphis, and Houston. This resulted in incremental line-haul transit times of 17 days for St. Paul, 11 days for Chicago, 7 days for St. Louis, 9 days for Cincinnati, 3 days for Memphis, and 2 days for Houston. The speed disadvantage of COB does not include reduced service frequency (biweekly or weekly), implicitly assuming shippers are able to schedule to

TABLE 8 COB BREAK-EVEN VOLUMES OF BOXES: WEEKLY SERVICE BETWEEN A SINGLE PORT AND NEW ORLEANS

Port	COB Rate[a]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[b]	COB Rate[c]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[d]
St. Paul[e]	520	35	48	360	51	70
Chicago[f]	290	51	71	190	78	108
St. Louis[g]	240	68	45	150	109	76
Cincinnati[g]	350	58	40	240	85	59
Memphis[g]	140	87	60	90	135	94
Houston[g]	140	101	70	90	157	109

[a]Volume-incentive rail rates less $110 per box, $50 per box savings to attract shippers, and $60 per box terminal costs.
[b]Based on one barge per trip.
[c]Estimated double-stack rail rate less $110 per box.
[d]If utilization is greater than 100 percent, service is not feasible.
[e]Biweekly service for 9-month navigation season.
[f]Biweekly service for 10½-month navigation season.
[g]Weekly service for 12-month navigation season.

SOURCE: Computed by LSU Ports and Waterways Institute.

TABLE 9 COB BREAK-EVEN VOLUMES OF BOXES: WEEKLY SERVICE BETWEEN TWO PORTS AND NEW ORLEANS

Port	COB Rate[a]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[b]	COB Rate[c]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[d]
St. Paul[e]	520	27	38	360	39	54
Chicago[f]	290	39	54	190	59	82
St. Louis[g]	240	55	38	150	89	61
Cincinnati[g]	350	49	34	240	72	50
Memphis[g]	140	64	45	90	101	70
Houston[g]	140	79	54	90	123	85

[a]Volume-incentive rail rates less $110 per box, $50 per box savings to attract shippers, and $60 per box terminal costs.
[b]Based on one barge per trip.
[c]Estimated double-stack rail rate less $110 per box.
[d]If utilization is greater than 100 percent, service is not feasible.
[e]Biweekly service for 9-month navigation season.
[f]Biweekly service for 10½-month navigation season.
[g]Weekly service for 12-month navigation season.
SOURCE: Computed by LSU Ports and Waterways Institute.

TABLE 10 COB BREAK-EVEN VOLUMES OF BOXES: WEEKLY SERVICE BETWEEN THREE PORTS AND NEW ORLEANS

Port	COB Rate[a]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[b]	COB Rate[c]	Break-Even Boxes per Week	Percentage Barge Utilization per Trip[d]
St. Paul[e]	520	24	34	360	35	49
Chicago[f]	290	35	48	190	53	74
St. Louis[g]	240	51	35	150	82	57
Cincinnati[g]	350	46	32	240	68	47
Memphis[g]	140	58	40	90	89	62
Houston[g]	140	72	50	90	81	78

[a]Volume-incentive rail rates less $110 per box, $50 per box savings to attract shippers, and $60 per box terminal costs.
[b]Based on one barge per trip.
[c]Estimated double-stack rail rate less $110 per box.
[d]If utilization is greater than 100 percent, service is not feasible.
[e]Biweekly service for 9-month navigation season.
[f]Biweekly service for 10½-month navigation season.
[g]Weekly service for 12-month navigation season.
SOURCE: Computed by LSU Ports and Waterways Institute.

meet COB sailings as they do to meet daily rail service. It was also assumed that the rate a COB service could charge would have to be decreased to cover increased carrying costs associated with only the differences in transit time, ignoring service frequency.

The results of the evaluation of transit time differentials, disregarding service frequency, are given in Table 11 for the three levels of terminal costs and for carrying costs of $10 and $20 per day. As should be evident from the table, carrying costs of $30 per day would totally negate any line-haul savings that might be achieved by a COB service. In most cases the COB rate would need to be less than $50 per box (and in some cases the rate would have to be negative) in order to attract cargo—a rate at which COB cannot be self-sustaining. As should be expected, the introduction of in-transit carrying costs substantially increases the break-even levels. Even in the case of $10 per box carrying costs, a cost that corresponds to empty container insurance and opportunity costs, break-even utilization rates greater than 50 percent are required. At $20 per box, a cost that would reflect relatively low-value merchandise, the feasibility of COB largely disappears. Except in the case of low terminal costs, the utilization rates required for a profitable service are probably not achievable on any sustained basis. As a result, it is concluded that implementation of COB service is not feasible except where there are significant volumes of very low-valued shipments that are not sensitive to transit time and service frequency.

ASSESSMENT OF COB POTENTIAL

The evaluation of the cost and operational aspects of Mississippi River COB services indicates that COB can be economically and technically feasible only if certain market conditions exist. For example, scheduled, weekly general towing service on the Mississippi River system would require a minimum of 3,000 to 6,000 boxes a year to break even, depending on vessel itinerary. The volume of boxes needed to break even is sensitive to the distances between interior ports and New Orleans. The further upriver, the lower the annual COB break-even threshold, approaching 3,000 boxes a year at St. Paul. As river distances increase, however, COB

TABLE 11 COB BREAK-EVEN LEVELS WITH IN-TRANSIT INVENTORY CARRYING COSTS

	$10 per Box Break-Even Boxes per Week	Carrying Cost Percentage Barge Utilization[a]	$20 per Box Break-Even Boxes per Week	Carrying Cost Percentage Barge Utilization[a]
Low Terminal Costs				
St. Paul	52	72	101	140
Chicago	82	114	210	293
St. Louis	96	67	164	114
Cincinnati	78	54	120	83
Memphis	110	77	152	106
Houston	118	82	142	98
Moderate Terminal Costs				
St. Paul	56	78	117	163
Chicago	95	132	327	455
St. Louis	113	78	218	151
Cincinnati	115	80	140	97
Memphis	143	99	221	153
Houston	149	103	189	131
High Terminal Costs				
St. Paul	61	84	140	194
Chicago	100	139	738	1,024
St. Louis	136	95	327	227
Cincinnati	197	67	170	118
Memphis	203	141	405	281
Houston	202	140	283	197

[a]If utilization is greater than 100 percent, service is not feasible.

becomes almost unacceptable for any time-sensitive traffic, because of the long transit times, variability of transit times, and winter closure of the waterways system.

Although the absolute number of containers necessary for profitable service appears to be relatively small, COB break-even volume is substantial in comparison with the current levels of container traffic moving by rail and truck between inland river cities and the Port of New Orleans. COB threshold break-even volumes would almost certainly necessitate diverting non-time-sensitive traffic away from other ports. Because almost all containerizable general cargo has already been diverted from break-bulk, except in lesser developed nations, COB would need to secure substantial commitments of relatively non-time-sensitive cargoes before a service could be feasibly initiated.

Implementation of new COB services entails significant risk in the absence of guaranteed, steady, balanced traffic flows. COB costs for line-haul, terminal, and overhead on a per unit basis are relatively constant over a wide range of volumes and vessel itineraries. Only the labor costs for loading and unloading have some variability. As a result, unless traffic commitments can be secured to underwrite the fixed costs of the service, a COB venture should be regarded as speculative. This assertion reflects both the analysis conducted in this study and the failures of recent COB endeavors.

The break-even projections for COB are quite sensitive to assumptions about terminal costs, rail rates, inventory costs, and overhead costs. Terminal costs will be a function of capital intensity, volume, and productivity. Terminal costs have assumed the use of nonunion labor or modified union manning levels. If union-scale wages and crew sizes were used, terminal costs would be almost doubled. For example, terminal costs at Memphis would be between $90 and $100 per box (lift-on or lift-off), and terminal costs at New Orleans would be about $125 per box (lift-on or lift-off). Conventional terminal costs of this magnitude would prohibit a COB venture.

Rail rates used in this analysis were for single shipments of 20-ft containers. The intermodal rates given in Tables 2 and 3 reflect a single 20-ft container tendered by one shipper on one bill-of-lading. The realities of the rail intermodal pricing structure allow shippers to tender two 20-ft containers on one bill-of-lading for slightly more than the price for a single 20-ft container. Therefore, unless individual COB shippers cannot aggregate pairs of 20-ft containers, the rail rates used for comparative analysis are approximately two times those that steamship lines or freight forwarders would incur for multiple shipments of 20-ft containers on one bill-of-lading. The result is that the data in Tables 2 and 3 represent the theoretical maximum rates for individual shippers of single 20-ft containers without any combination of containers by freight forwarders or shipper consolidators.

Unless shipments have zero or quite low time sensitivity, a weekly COB service, which potentially could increase average transit times between 10 and 20 days for midwest ports, will not be economically attractive because of high inventory costs. All indications are that time-sensitive cargoes, which comprise the bulk of containerized cargoes, would be unable to use COB and derive any significant transportation savings. Therefore, COB would have to attract neobulk and relatively low-value container shipments, such as repositioning empty marine containers, in order to be successful.

COB break-even projections are also sensitive to assumptions about drayage expenses and chassis utilization. If drayage costs between COB interior river terminals are significantly less than those of rail because of shorter distances or less congestion, COB break-even thresholds will be lower. If container chassis utilization is reduced, however, COB costs will increase. Of particular importance in this regard is an operational structure that will attract streamship lines as active promoters of COB services.

Nonetheless, COB has significant attractive features. The cost analysis indicates that a competitive pricing structure is possible only if sufficient volume of single 20-ft containers not subject to consolidation for lower rail rates exists and if productivity is high. Depressed conditions in the towing industry, characterized by an oversupply of equipment, should enable prospective COB operators to lease all equipment at nominal rates. Moreover, towing services can be negotiated at levels that are significantly lower than published tariff rates. With small terminal crew sizes, COB dock-to-dock line-haul and transshipment costs could be competitive with rail for non-time-sensitive cargoes.

Implementation of COB service requires several steps if the service is to be profitable for the inland and ocean carriers, and sufficiently cost competitive to attract the necessary volumes to achieve high levels of equipment utilization and terminal productivity. The successful examples of COB indicate that a signigicant amount of market research was conducted before implementation. This research indicated the levels of cargoes that might be available as well as the pricing and quality-of-service that COB would need to provide to be a viable alternative to existing intermodal shipment patterns. Perhaps as important as the market research is the ability to transform the market information into commitments on the part of shippers, ports, and other carriers. These commitments provide the base cargoes that underwrite initial COB services.

There must also be a commitment to structure COB as an intermodal service. This means that COB must provide an intermodal rate and service package beyond a towing charge. By definition this requires that, in addition to towing companies, other transportation entities be involved in developing an intermodal COB service. COB potential will be limited if the service is marketed as a dock-to-dock, for-hire towing alternative to land-based line-haul services that are part of an intermodal distribution system. The integration of railroads, ports, and steamship lines, including trucking and drayage, is resulting in a new one-stop shopping dimension to intermodalism.

COB must be conceived and executed as part of a through intermodal service, not a fragmented alternative to one component of an integrated intermodal package. The service cannot simply be integrated into existing towing operations. It will require adaptation of inland towing operations to intermodal operations. COB must be a scheduled, reliable service if it is to be operationally competitive with the land-based modes. Reliability includes not only the towing operation but also container and chassis availability, yard security, loss and damage control, and other physical distribution characteristics. Without integration of COB into an intermodal service, the lack of necessary operational features, such as container pools and high productivity levels at inland terminals, will effectively block implementation of new COB services.

To be successful, COB must be structured as a distribution package, and barge and towing companies must be able to effectively market an intermodal COB service. Intermodal COB can be an extension of a shipping line, a consortium of shipping lines, a port agency, or a shipper cooperative. The formal organization of COB is not particularly important, however; unless COB has a through rate and intermodal service package that can be incorporated into ocean service tariffs and service contracts, COB has relatively limited prospects and potential. In spite of impressive potential savings for long-distance cargoes and competitive rates for short-haul cargoes, COB must reduce total distribution costs to present a successful alternative to emerging intermodal systems.

Publication of this paper sponsored by Committee on Ports and Waterways.

Privatization Is More than Contracting Out

FRANK W. DAVIS, JR., W. DAVID SMITH, AND WILLIAM J. HEWA

Privatization is rapidly becoming a popular option for increasing the efficiency and effectiveness of service provision. Unfortunately, too often privatization is equated with contracting out. In this paper privatization is distinguished from contracting out, when privatization is necessary is discussed, and a four-step approach to implementing privatization is suggested. An understanding of privatization concepts requires an analysis of the services to be delivered and the contracting environment. If agencies blindly apply traditional contract procedures in the wrong environment, the result is frequently the "publicization of private service," which has the worst characteristics of both the public and the private sector.

Some people view privatization as another fad like urban renewal; others consider it a philosophic stance. Privatization is more than a simple fad or philosophy; it is, or should be, an expansion of traditional economic and management thought. Nobel laureate James Buchanan initiated much of the interest in privatization when he emphasized that government financing of an activity is not the same as government production of a service. These are two distinct and separate activities. This is clear from his statement that "governmental financing of goods and services must be divorced from direct governmental provision or production of these goods and services" (*1*). Many individuals have taken this to mean that the private sector should provide all goods and services because it is more cost-effective than government. This is too simplistic a view.

Successful privatization requires a thorough understanding of the contract marketplace, for this is where the interaction between the government buyer and the private-sector provider takes place. Successful privatization is more than issuing an invitation for bid (IFB) and awarding a contract. Successful privatization requires a thorough understanding of the contract environment, the nature of the service desired, and the response to the various soliciting methods. The contract marketplace is just as complicated as the traditional marketplace around which many disciplines such as microeconomics and marketing have developed.

FRAGMENTATION OF ACADEMIC DISCIPLINES

The study of economics and management has been fragmented into many different disciplines, each studying a different sector of the economy within carefully defined parameters. Unfortunately, none of these disciplines specifically addresses the contract marketplace or the interaction of public, private, and nonprofit sectors. Examples of the most familiar disciplines follow.

Department of Marketing, Logistics, and Transportation, The University of Tennessee, Knoxville, Tenn. 37996-0700.

1. Microeconomics studies for-profit firms producing products for sale in the marketplace. The public sector and the nonprofit sector frequently do not fit the assumptions of this discipline.
2. Public finance traditionally studied the collection of funds to provide public goods in a command economy in which authority is derived from the electoral process. Only recently have specialists in public finance discussed the need to separate the production of public good from the financing of public goods. There are fewer alternatives for management when the public sector procures goods and services.
3. Regulatory economic theory addresses firms that require such an extensive network of facilities to deliver services that they are in essence natural monopolies. This discipline provides the basis for regulating firms such as railroads, distributors of electrical power, bus lines, and communications.
4. Business administration courses address the management of for-profit firms producing products for the marketplace. The accounting discipline does recognize the other sectors but has three different sets of accounting practices: one for for-profit firms producing products, one for government organizations, and one for nonprofit organizations. A case can be made that there is still a fourth approach for public utilities. Each practice is substantially different in many ways, such as the handling of depreciation and budgeting.
5. Public administration focuses on command management of public organizations that receive a mandate (authority) from a legislative body and are responsible for carrying it out according to the guidelines given.

Generally missing is a comprehensive approach to understanding the dynamic interaction among the various sectors (private, public, nonprofit), especially when they provide services rather than produce products. This fragmentation is not new; it can be traced from the very roots of economic and business disciplines.

Privatization is an evolving managerial process operating without the benefit of a well-understood body of management principles such as business administration or public administration. What is needed is a set of principles to guide public administrators as they procure goods and services from nonprofit and for-profit providers. These principles need to answer questions such as

1. What actions encourage more effective delivery of the desired services?
2. What factors control costs?
3. Is there more than one way to contract for goods and services?
4. When should the various contracting methods be used?

5. How can the contracting environment be controlled to reduce the risk and hassle of contractual arrangements?

If privatization is to evolve successfully from a theoretical concept to a widely used mechanism for controlling costs and improving the fit between agency needs and contractor-provided benefits, the way the contract marketplace works must be better understood.

WHAT IS PRIVATIZATION?

Kolderie (2), building on Buchanan's definition, suggests a model for classifying privatization efforts. According to this model, service delivery consists of two parts: provision (initiation and funding) and production. Provision "is the policy decision actually to provide a good or service." Production "is the administrative action to produce that good or service." This classification suggests four ways that services might be provided:

- Case 1: Government does both when the legislature establishes and funds a public organization to provide the service so that neither function is private.
- Case 2: Provision is public but production is private when government hires a contractor to provide a needed service such as road construction.
- Case 3: Government produces the service but the private sector pays for it as would be the case when a builder contracts with the city to provide policemen to control traffic where large trucks enter the highway.
- Case 4: Both provision and production are private as would be the case if the contractor hired a private security service to control traffic.

These four cases can be displayed graphically as shown in Figure 1.

Production \ Provision of Service	Government	Private
Government	Case 1	Case 2
Private	Case 3	Case 4

FIGURE 1 Provision of products and services.

This classification system identifies several approaches to privatization:

- Case 1 or Case 2 to Case 4 privatization. Government-provided and government-produced activities such as fire protection, ambulance service, or trash pickup may be converted to totally private activities. This is typically done when the private sector rises to provide services that the public sector ceases to provide. Frequently this shift is accompanied by the regulation of the emerging private-sector industry as in the case of ambulance services or private fire departments.
- Case 3 to Case 4 privatization. Governments have been selling off Case 3–type government-owned enterprises such as British Steel, British Air, and British Telecom in England and Conrail in the United States. This usually occurs because the private sector is already buying the products or services and privatization is simply a matter of finding a willing buyer for the organization. Exceptions occur when the operation is heavily subsidized and shows no potential for profitability.
- Case 1 to Case 3 privatization. Government activities such as parks and recreation facilities may be partly privatized by charging those who use the service or facility. This is especially appropriate when government is reluctant to give up title to national properties such as parks and recreational facilities.
- Case 1 to Case 2 privatization. In the United States the word "privatization" is typically used to describe Case 1 to Case 2 privatization. In this case the production function that was formally performed by government is contracted out to a private operator. The private operator simply produces the product or service but looks to government as the source of revenue. The word "privatization" as used in this paper refers to Case 1 to Case 2 privatization.

Unfortunately, it has been too easy to equate contracting out with privatization. When this is done it often creates a system that is less cost-effective, less responsive, and more bureaucratic than in-house provision of public services. If privatization is to be effective, it must be based on a clear understanding of the process required to privatize successfully and a recognition of the economic principles that lower cost or improve effectiveness of delivery, or both. Success does not occur because private organizations are more responsive or efficient than public agencies. Success occurs when the service provider's success is dependent on improving the effectiveness of the service provided. Cost savings occur when the operating environment rewards providers for controlling costs. Thus successful privatization must be based on the use of correct procurement procedures and principles. The purpose of this research is to identify these procedures and principles and indicate the conditions that allow successful privatization.

CONTRACTING OUT IS NOT SYNONYMOUS WITH PRIVATIZATION

In the private sector, when markets are working effectively, contracts are typically implied, not written, especially when relationships are continuous over an extended period of time. The bread producer delivering bread to a grocery store will typically do so on an implied contract. The customer buying bread will do so on an implied contract. Under the implied contract the buyer agrees to pay the sum agreed on, and the seller agrees that there is no hidden defect, fraud, or misrepresentation. If the product or service is unsatisfactory it is made right. "Satisfaction is guaranteed" not by court action but by keeping the value of regular, repeat future business always worth more than the value of any current transaction. Under this arrangement, the bread manufacturer does whatever is necessary to keep the merchant happy and the manufacturer's brand on the shelf. The merchant's self-interest requires keeping the customer happy and coming back to buy more bread. Thus, in the market allocation world of day-to-day business, there are few written contracts.

Written contracts have an entirely different objective than does privatization. The principle of written contracts began in 1677 when the English Parliament prohibited bringing suit for fraud more than a year after an agreement was made unless the agreement was written (*3*). Thus written contracts were a method of spelling out all of the conditions of long-term agreements as understood by both parties at the time of agreement. This was done to prevent the confusion that occurred in court when the testimony of the two parties, polarized by the emotions of conflict and time, provided no objective method of determining intent at the time the agreement was made.

Written contracts are required by law when the agreement

1. Cannot be carried out fully in less than 1 year;
2. Is made in consideration of marriage;
3. Is for the sale of land;
4. Is to serve as an executor or administrator of an estate;
5. Is for suretyship to be responsible for the debt of another person (*4*).

In practice written contracts are typically used only when

1. The desired benefit is clearly understood by both parties to the extent that all details can be stated explicitly in advance. If the need is still so general that it can only be stated in terms of "high quality," "satisfactory service," "fair prices," and "to be provided when needed," it is difficult to write a contract.
2. The agreement, such as a conditional sales contract or a lease, is for an extended time period.
3. A large purchase is involved and there may be extreme risk to either or both parties, as with the purchase of a home or business.
4. Delivery is to occur in the future as in the case of constructing a home or payment of insurance.

Attorneys who write contracts are trained to write comprehensive, consolidated contracts. Contracts are considered comprehensive when they cover all possible future eventualities. Contracts are considered consolidated when all possible aspects of the agreement are covered. A comprehensive contract would cover not only the price but also such items as method to be used, conflict resolution, performance criteria, all possible conditions for nonperformance, and contract changes.

The purpose of contracts is to create certainty and stability. A contact is appropriate when the intent of both sides is explicit (certain) and the agreement is for an extended period of time (stability). The objective of privatization is to increase cost-effective delivery. Privatization encourages flexibility, multiple suppliers, competition, improved fit, and reduced hassle. Contracting typically tends to reduce flexibility, reduce variation in service provided, and eliminate competition. A marriage contract, for example, does not encourage the husband or wife to seek new competitors but to make a long-term commitment that excludes competition regardless of future situations.

An example is the Corrections Corporation of America's offer to pay the state of Tennessee $250 million for a 99-year lease on the entire Tennessee prison system. Like a marriage contract, this offer was designed to establish a long-term operating agreement with the state for a period of 99 years without fear of competition. Consider the situation if the state had entered into this agreement. If the state later sought a "divorce" within the 99-year period, a massive "property" suit could follow before the state would be able to operate its prisons again. If the state became dissatisfied with the operating arrangement, the court might require "alimony payment" to finalize the "divorce."

Agencies should remember that the contracting process can cancel out all of the hoped-for benefits of privatization and even eliminate much of the flexibility currently available under agency production of service. When this is done, the problems that result are not the results of privatization; they are caused by the way privatization has been implemented. Successful privatization requires that each step in the process be planned and executed with a thorough understanding of the consequence of not following each principle.

PRIVATIZATION PROCEDURES

The four steps to privatization are

1. Defining the need and selecting a procurement strategy,
2. Soliciting qualified providers,
3. Developing contracts that control but do not increase risks, and
4. Monitoring contracts to improve the fit between desired and delivered benefits.

Each of these four steps must be based on an understanding of the service needed and the contracting environment. Traditional procurement procedures function best when a need is recognized, the planning process can be followed to determine the best way of meeting the need, and procurement is completed after the need is recognized. But what happens when the need cannot be defined before procurement?

NEED TO DEFINE FUTURE NEEDS THAT ARE NOT YET KNOWN

The first step in privatization is defining the need to be privatized. The more detailed the definition, the more exact the bidder can be in responding. On the other hand, the more vague the requirements, the more the bidder needs to "pad" the bid to cover unexpected contingencies. If the service definition is too vague, the bidder cannot be responsive. (A responsive bidder is one who meets all of the qualifications and specifications.)

Consider the case of providers of transportation for the elderly and handicapped attempting to define specialized service needs. The total number of trips for the contract period is not known. The origin and destination of trips are not known and will only be determined after the individuals who need to travel and their trip purposes are known. The trip schedule will not be known until the purpose of the trip, such as a doctor's appointment, is known. Because these details will often not be known until the day of the trip, it is impossible to define the service for purposes of competitive bidding.

Attempts to provide this type of service using competitive bidding force an agency to define something that cannot be known. Consequently, to meet procedural requirements, the agency defines the way it thinks the contractor should organize. The resulting organization ends up being similar to what would have been provided by the agency itself. The number of vehicles to be available, the dispatching system, vehicle specifications, driver qualifications, and management style are specified. When the procurement process forces an agency into this type of contracting, it tends to eliminate most of the options that enable private management to be more cost-effective. In essence, the procurement process forces the agency to procure equipment and an organization instead of service. This approach can "publicize private providers" rather than privatize the provision of public services.

NEED TO DEFINE TECHNICAL SPECIFICATIONS THAT ARE NOT YET UNDERSTOOD

Often an agency is faced with procuring an unfamiliar product or service. When buying unfamiliar goods such as a new computer, a special-use vehicle, or a software package for the first time, staff are often unsure of the best options. Certainly they are unlikely to be knowledgeable about highly specific dimensions and technical specifications.

Having heard "horror stories" of van suppliers delivering vehicles without seats because they were not included in the specifications for the 15-passenger van, the agency staff want to make sure that all details are covered adequately. To do this they will look at the product provided by a vendor whom they trust. This trust may come from personal experience, from friendship, or from the halo effect of a firm that is large or highly publicized, or both. (How often has it been said that no data processing manager has ever been criticized for deciding on IBM because so many others have made the same choice?) If the staff are comfortable with the vendor's salesperson, they frequently ask the vendor to supply a list of specifications because they recognize that the vendor is more knowledgeable about the technical aspects of what the agency needs than are the staff. Not surprisingly, the vendor-supplied specifications may unduly emphasize those specifications that will preclude competition, whether or not the requirements are relevant. Thus competitive bidding becomes sole-source acquisition—a potential problem.

SOLICITATION

The federal government has a long history of procuring products and services. Traditionally procurement was for military supplies; for products and services for use in government operations; and for the construction of facilities such as buildings, roads, and bridges. Beginning with the New Deal programs, government began to procure services that were offered directly to the public. These programs include public transportation, social service programs, and care of the elderly. As these programs began to evolve, procurement methods received new attention.

The systematized organization and formulation of rules and methods of procurement began with the Federal Procurement Regulations (FPR) codified by the Administrator of the General Services Administration (GSA) pursuant to the Federal Property Administration Act of 1949. For more than 30 years the FPR contained the chief governing rules for a myriad of executive agencies that were procuring more and more goods and services for particular groups of clients. In 1984 attention was finally turned to developing a unified single set of rules for federal procurement that would allow all agencies to procure at a cost-effective rate, preferably by competitive bidding. This desire was expressed by Congress in the Competition in Contracting Act of 1984. At the same time, the Federal Acquisition Regulation (FAR) system (codified at 48 CFR § 1-15) became effective as the governing rules and regulations for federal procurement. Although concise and uniform, the FAR represents literally volumes of procurement rules.

The FAR may be intimidating reading, but it does systematically prescribe the procedures that should be followed to contract for a product or service. Unfortunately, the FAR is organized by procurement type. This organization and the size of the FAR virtually assure that readers will simply use the index to look up a specific procedural question without considering the document in its totality and understanding the more basic concept of when each method should be used.

VENDOR SELECTION METHODS LISTED IN THE FAR

The FAR not only includes a list of vendor selection methods but also presents them in an order of preference. The methods are

1. Acquisition through other agencies if they have usable surplus;
2. Procurement lists of goods and services available from institutions employing blind or severely handicapped persons or federal prison industries;
3. Supply schedules indicating goods and services available from various vendors at previously agreed on prices and conditions;
4. Invitations for bids, including requirements contracts, bids from qualification lists, bids with samples, and two-step process;
5. Requests for proposals (these may be preceded by requests for information);
6. Sole-source contracts; and
7. Unsolicited proposals.

Acquisition Through Other Agencies

Highest priority for acquisition is given to redistribution among agencies and solicitation from protected sources. The top priority is for agencies to satisfy their needs from agency inventories (in-house) or excess inventories in other agencies. This is an effort to encourage agencies to use existing supplies, especially when other agencies have a surplus.

Protected Sources

The second priority, if the needed goods or services are not found in inventory, is to select a provider from the procurement lists of goods and services available from institutions employing blind and other severely handicapped persons or federal prison industries.

Supply Schedules

If the required items or services are not available from other agencies or protected sources, agencies should turn to the supply schedules negotiated by the GSA. The supply schedules provide federal agencies with a simplified process for obtaining commonly used supplies and services while still providing the economies of volume buying. The schedules are basically open-ended supply contracts negotiated on a periodic basis by GSA personnel. The individual agencies do not have to contract for items on the supply schedule. Instead, they can order and receive goods through a process similar to mail order purchasing via catalogs. The supply schedules are not limited to products; they include a wide variety of services such as hotel rooms, car reservations, and airline fares.

The strength of the supply schedules is their ability to supply a diversity of products and services with short lead times and a minimum of contracting effort. Supply schedules can be updated quickly and frequently. The GSA is continually adding new providers to the list. Any time a new product or service is introduced, an addition can be made to the supply schedule to handle this new variation. If a need occurs only intermittently, the supply is still there in the form of an open-ended contract at a stated price for a stated period of time. When the product or service is needed the providers deal directly with the agency that needs the supply (48 CFR § 38).

Invitations for Bids

When a need cannot be satisfied from the supply schedule, a provider may be found through competitive bidding. This typically occurs when

1. The purchase is of a nonstandard good or service,
2. Delivery conditions and terms are somewhat nonstandard, or
3. The purchase is so large that providers can be expected to quote an even lower price in a bid.

The competitive bidding process requires that the agency completely define the need so the specifications can be incorporated into the IFB. Next, a list of potential bidders must be compiled so they can be sent the IFB. Because the need has been totally defined, the only remaining decision is the selection of the provider and this is done on the basis of price. Thus the contract is automatically awarded to the lowest responsive bidder as of the submission deadline.

Managing the solicitation process requires the careful balancing of two factors. Theoretically, the larger the number of responsive bidders, the greater the competition and the lower the price. Therefore the list of potential providers must be inclusive enough to allow a number of responsive bidders. Potential providers can ask to be placed on bidders' lists or agencies may add providers on the basis of past experience or recommendations from other agencies. On the other hand, attention needs to be given to the bidders' list to make sure that only qualified bidders are allowed to bid. Unless this is done, an unqualified bidder may submit the lowest bid and win the contract.

Not surprisingly, sealed competitive bidding is unpopular among contracting officers because they have no subjective control over the award. If conditions or needs change, the chances of having to renegotiate, modify, or terminate the contract after it has been awarded can be relatively high (5). For this reason, sealed bidding should only be employed when the goods or services required are so well defined that any responsive bidder could provide them successfully.

There are several key variants of the IFB procurement process. The first of these, the requirements contract, allows postprocurement variation in definition. The last three are methods for qualifying respondents.

Requirements Contracts

Often a contracting officer will need products or services that are well defined except for the quantity needed or the exact delivery dates, or both. For example, for a construction site the specific type of cement needed may be known but not the exact quantity needed or the specific days that the cement will be poured. In such instances indefinite-delivery or indefinite-quantity contracts (commonly referred to as requirements contracts) are the best procurement methods. Usually the agency's obligation under these contracts is stated as some minimum quantity to be purchased over a maximum time period for delivery. Requirements contracts are awarded under competitive sealed bidding and in some cases under requests for proposals.

IFBs with Qualification Lists

When there is wide variation among the ability of various providers to successfully deliver the desired products or services, the procurement officer will attempt to qualify those that are allowed to bid. The agency does not want to risk the simple IFB procedure because inexperienced providers may attempt to "buy in" by bidding low (5). For example, an agency may require an architectural firm specializing in free-span bridge construction consultation services to prove that it has successfully completed bridge projects of similar size and complexity. The contracting officer knows that although there are numerous potential providers, many are relatively inexperienced. To avoid the danger of an inexperienced firm buying in, an officer may employ the IFB with qualification lists method. Bids will be accepted only from providers on the list (or who meet the requirements). This method encourages competition but only among contractors who have a high probability of successful delivery.

IFBs with Bid Samples

Another method of making sure that the delivered product effectively meets the need is the IFB with bid samples required. The samples may either show design improvements or serve as a portfolio of previous work. For instance, a wildlife and fisheries agency may need a customized inventory software package. Many providers may be able to write such a program, but there may be a big difference in user friendliness, completion time, and documentation. IFBs with bid samples, often referred to as first-article testing, allow the agency to be sure the potential providers can fill the need but still allow for competition. Each potential bidder must present a sample for testing or evaluation before being allowed to place a bid. Ultimately the lowest bidder is automatically chosen, as in the regular IFB process. Although the bid sample process allows the procuring officer to "kick the tires and blow the horn" before making a purchasing decision, it requires lengthy periods for preparation and testing on the part of the providers (*5*). The added costs may tend to drive potential bidders away unless the item is a relatively standard off-the-shelf item.

Two-Step IFB

Another variation on the simple IFB process is useful when the specifications and definition of goods and services are somewhat vague but the goods or services are potentially obtainable from many sources. An example of this would be contracting out for handicapped transit services when the exact customized facilities needed are not yet known by the contracting officer. Several providers may be capable of transporting the clients, but just how is a question. The two-step IFB process begins with potential bidders submitting technical and administrative proposals. On the basis of these proposals the contracting officer eliminates those providers that he or she believes are not qualified to bid on the contract. This subjective process allows the officer enormous flexibility in eliminating less-than-desirable providers. Again, the contracting officer does not have any choice in selecting the ultimate contract winner—the low bidder among the firms allowed to bid—and the additional screening will require additional planning and effort on the part of the procurement officer.

Requests for Proposals

Often the contracting officer really cannot define the best way of satisfying the agency's need. This lack of definition may result from the agency's lack of technical knowledge or simply because the need is unique or new to the people involved. For example, an agency may need to research a problem that it is facing for the first time. When such a situation exists, the agency will typically use a request for information or request for proposals (RFP) so the agency can evaluate the proposers' understanding of the problem and approach to solving it. This way the evaluators can increase the chance of successful contract completion. According to one source, 80 percent of federal procurement monies in 1978 were awarded using RFPs (*5*).

RFPs have the added benefit of providing the agency staff with many suggestions that can be used to further define an approach to meeting the need before the final award is made. The final award need not conform to the approach proposed by any single provider; it may be a composite that the agency believes will be most effective after considering all approaches submitted. Some proposers are reluctant to submit ideas if the agency is likely to integrate them into the final definition and then award the project to another contractor. RFPs do not restrict the award to the lowest bidder or even the proposer with the most creative ideas; RFPs allow agencies to select the provider they believe will produce the best results.

Sole-Source Contracting

When goods and services can only be responsibly delivered by a single provider, sole-source and unsolicited proposal procurement methods may be employed. The FAR authorizes agencies, with appropriate justification, to negotiate with a single provider if no other responsible providers exist and no substitutes are acceptable.

Unsolicited Proposals

The unsolicited proposal method of procurement is often employed by procurement officers when a research entity devises a new product or method of providing a service that will enhance an agency's ability to serve its clients. These entities, which may be institutional or private, are given research contracts to develop the good or the method of delivering the service in question.

APPLICATION OF SOLICITATION METHOD

Too often agencies believe they are constrained by solicitation options. Success is too often viewed as being able to "get a contract through." Likewise, procurement is too often viewed as a gatekeeper, an inhibiting factor that limits the agency's service to its clients.

This need to "beat the system" often indicates that the individual does not understand the procurement process, especially not the way the solicitation process can help create an environment that reduces the hassle and increases the effectiveness of procurement. To understand how the procurement process can be used to create a more favorable contracting environment, it is necessary to first understand how each procurement method can and should be used.

The competitive bidding process works well for a single, totally defined need. The need has been recognized and alternative approaches have been explored before the procurement process is begun. Because everything has been totally defined, only two questions remain to be answered in the procurement process: Is the potential provider able to meet the specifications? What are the costs of meeting the specifications? The variations in the bidding process (qualification list, bid samples) are simply efforts to qualify the bidders to ensure that they are able to meet the specifications.

RFPs are used for a single purchase when the need is totally defined by the agency but many alternate approaches may exist for delivering the defined benefits. RFPs allow providers to suggest the methods that they would use and explain their understanding of the problem. The solicitation process has three parts: (a) deciding which delivery approach appears most promising, (b) determining the ability of the provider to do what is proposed, and (c) determining the cost of delivering the needed benefits using the approach proposed.

A requirements contract is used when the agency knows exactly what is needed but not the time of delivery or the expected quantity. For example, the GSA has a requirements contract for airline services between approximately 1,951 city pairs. The rates and carriers are agreed on, but the number of trips made and the date and time of each trip are left open.

Because requirements contracts are generally for a specific product or service, they are only negotiated when there is an existing and continuing need for the predefined product or service. When there are many diverse needs, a supply schedule is used. The supply schedule is a list of all requirements contracts for certain categories of products or services. For example, the GSA has a supply schedule for microcomputers and related software and repair services. The supply schedule was not negotiated in response to an actual need; GSA recognized the general need for products and services such as office supplies, computers, travel services, and other routinely needed benefits. Instead of conducting a new solicitation each time another purchase is proposed, GSA asks all qualified providers to submit specifications and prices. These conditions and prices are reviewed by the contracting officers. If the price appears to be reasonable and the discounts are significant (government should receive the largest discount because it is the largest purchaser), the supplier is included in the supply schedule. A given vendor may have many different items included in the supply schedule. Likewise, many different vendors of similar products or services are included in the supply schedule. Because there are so many different vendors on the supply schedule, an agency has the option of selecting the vendor whose product or service is most effective in meeting the agency's specific needs.

The bidding method, the proposal method, and the requirements method are predicated on several assumptions:

1. There is adequate lead time between recognition of need and procurement to allow the solicitation process to work. Specifications can be developed and potential providers can be notified and given adequate time to respond.
2. Solicitation is for a single purchase.
3. There are many potential providers willing and able to provide what is needed.
4. The price is determined primarily by the size of the purchase and not by the timing of the purchase.
5. The procurement process will not disrupt or change the marketplace (i.e., create a monopolistic situation).

Unfortunately, these theoretical conditions do not hold for many services purchased by local governmental bodies. Some of the reasons for this follow.

1. Many services must be purchased before the need is defined.
2. Instead of independently soliciting for each need, agencies attempt to collectively procure to satisfy many different needs at one time. Transportation for the elderly and handicapped is a prime example. Each trip is a separate need just as each construction contract project is a different need. When the agency attempts to group needs to "make procurement easier," it tends to negotiate noncompetitive "marriage" contracts.
3. When a community is in the transition stage between system and network, potential providers do not exist.
4. Because service providers cannot inventory their capacity, prices may fluctuate more widely by time of purchase than size of purchase. For example, where a network of private school bus operators exists, inexpensive transportation can be obtained when the provider is not transporting school children. If service is needed at the same time as regular school bus runs, then the cost will be quite high because added capacity will have to be scheduled.
5. When an agency awards an all-or-nothing "marriage" contract for a category of service over an extended time period, such as transportation for the elderly and handicapped for a year, it tends to monopolize the marketplace.

When the solicitation environment does not meet the five conditions required for competitive bidding and RFP solicitations, modifications must be made to meet the intent of the Competition in Contracting Act of 1984.

CONCLUSION AND SUMMARY

Privatization is not a simple process. It is complex and has many different dimensions depending on when the agency can define the need and the nature of the contracting environment. Most of the criticism of privatization occurs when procurement officers and agency attorneys incorrectly use "approved methods" that are simply wrong for the conditions. The authors are currently completing a study that addresses several basic questions: How should a need be evaluated to determine the most appropriate way of meeting it? When should the solution consist of a system and when should a network be used? What is the basis for determining the most appropriate solicitation method? Is a contract needed and, if so, when should various clauses be included? What is the role of the contract monitor? Is it the same when the contract is for a system, for a network, or for something that is not totally defined? It appears from this analysis that the public procurement market under privatization is fully as complex as the marketplace for goods. Just as the disciplines of economics and business consider many different markets, so they must begin to address the many different procurement markets.

ACKNOWLEDGMENT

The authors wish to thank the Urban Mass Transportation Administration for its funding that made this paper possible. Funding was provided by the Office of University Research and Training.

REFERENCES

1. J. M. Buchanan. Why Does Government Grow? In *Bureaucrats: The Source of Government Growth* (T. E. Borcherding, ed.), Duke University Press, Durham, N.C., 1977.
2. T. Kolderie. The Two Different Concepts of Privatization. *Public Administration Review*, July–Aug. 1986, pp. 285–291.
3. J. Calamcli and J. Periollo. *Contracts*. West Publishing Co., St. Paul, Minn., 1977.
4. S. Emanuel and S. Knowles. *Contracts*, 2nd ed. Emanuel Law Outlines, Inc., New Rochelle, N.Y., 1984.
5. J. Robertson. *Selling to the Federal Government*. McGraw-Hill Book Company, New York, 1979.

Publication of this paper sponsored by Committee on Manpower Management and Productivity.

Successfully Establishing a Strategic Planning Process

MARK P. HOWARD

The Metropolitan Transportation Authority (MTA) of the state of New York has established a multiyear strategic planning program, called the Strategic Planning Initiative (SPI), in order to respond effectively to the changing character of ridership demand in the context of available financial resources and to the threats and opportunities that arise as the region evolves. The SPI represents an ambitious attempt to examine the strengths and weaknesses of the services provided by the operating agencies under the MTA umbrella. This assessment, outlining the way the systems are now being used by passengers, what it costs to run them, what changes in transit demand are likely during the next 20 years, how to regain lost ridership, and what kinds of responses could be made to shortfalls in capital funding, is aimed at helping the key decision makers at MTA, and in local and state government, to understand the implications of the choices being made today as well as some alternative ways of looking at issues and options. The methodology developed for the SPI Work Program for Year One includes three main elements: (a) development of three financial scenarios assessing the trade-offs that would be necessary under reduced capital funding; (b) data collection and analysis including ridership data and operating cost by line models; and (c) development of service guidelines to determine when, where, and how much service should be provided to the public.

The Metropolitan Transportation Authority (MTA) of the state of New York has established a multiyear strategic planning program, called the Strategic Planning Initiative (SPI), in order to respond effectively to the changing character of ridership demand, in the context of available financial resources, and to the threats and opportunities that arise as the region evolves. This 3-year process, begun in the spring of 1985, is scheduled to end in 1988.

The MTA is a multipurpose state agency responsible for operating mass transportation services and facilities in the New York metropolitan region. The agencies governed by the MTA Board of Directors include the New York City Transit Authority (TA), Long Island Rail Road (LIRR), Metro-North Commuter Railroad (MNCR), Metropolitan Suburban Bus Authority (MSBA), and Triborough Bridge and Tunnel Authority (TBTA).

The SPI represents an ambitious attempt to examine the strengths and weaknesses of the services provided by the operating agencies under the MTA umbrella. This assessment, outlining the way the systems are now being used by passengers, what it costs to run them, what changes in transit demand are likely during the next 20 years, how to regain lost ridership, and what kinds of responses could be made to shortfalls in capital funding, is aimed at helping the chairman and members of the MTA Board, and local and state officials, to understand the implications of the choices being made today as well as some alternative ways of looking at issues and options.

The architects of the SPI also decided that it was necessary to make a compelling statement about the essential nature of the New York metropolitan area's mass transit resources. Two-thirds of the daily trips to the Manhattan central business district (CBD), for example, are made by public transportation. Nevertheless, highway congestion has been increasing throughout the region, and transit's share of the transportation market is slipping. Outside the Manhattan core, changes in demographic and employment patterns are working against traditional forms of transit, as jobs and people disperse into the lower-density suburbs and exurbs of New York, Connecticut, and New Jersey.

In addition to the challenges posed by changing demographic and employment patterns, there is concern about a continued loss of ridership, especially on the New York subway, resulting from rising fares, poor service, a degrading public environment, and fear of crime. The MTA's massive 5-year capital program for 1982–1986 is beginning to relieve some of these problems, as are management actions at each of the MTA agencies, but it is clear that additional actions to attract and maintain transit riders will be necessary.

The methodology developed for the SPI Work Program for Year One includes three main elements: (a) the development of three financial scenarios, assessing the trade-offs that would be necessary under reduced capital funding; (b) data collection and analysis efforts, including ridership data and operating cost by line models; and (c) development of service guidelines to determine when, where, and how much service should be provided to the public.

STRATEGIC PLANNING IN CONTEXT

Service Crisis

There were many reasons for the poor condition of public transit in the New York region, but the most obvious problem was money. The MTA was established in 1968 amid sweeping plans for new construction and expansion of the system. The New York City TA was to construct a network of new routes in the late 1960s and 1970s, including a Second Avenue subway on the East Side of Manhattan as well as a new subway express bypass and 63rd Street tunnel to relieve overcrowding in the

Metropolitan Transportation Authority, 347 Madison Avenue, New York, N.Y. 10017-3706.

fast-growing Queens-to-Midtown Manhattan travel corridor. The LIRR was to gain access to the East Midtown area through a new Third Avenue terminal, and also through the 63rd Street tunnel. The issue of reconstructing existing facilities was largely ignored.

By the mid-1970s, however, the bitter realities of the New York fiscal crisis had frustrated the dreams of expansion and had exposed major problems in the way the system was operated. Deferred maintenance of equipment and facilities, combined with a continuing capital funding shortfall for the rehabilitation of the existing physical plant, led to sharp increases in service disruptions. Ridership plummeted, and as inflation and labor cost increases forced operating costs up, the scarcity of operating subsidies led to hard choices between fare increases and service cutbacks. Factors beyond the control of the operators of the system led to even more ridership losses. The New York City employment picture changed, so that fewer people, working at different types of jobs, were traveling to the Manhattan CBD. Fear of crime. and the perception that the New York City subway system was an unsafe environment, also contributed to the erosion of the ridership base. Graffiti, combined with the decrepit condition of equipment and stations, presented an image of a system out of control. The rapid transit system was reeling from an unprecedented rash of derailments, fires, and service disruptions. On the bus system in New York City ridership was expected to continue to drop, in part as a result of obsolete service patterns.

In late 1979 a new MTA Chairman, Richard Ravitch, was appointed by Governor Hugh L. Carey. Ravitch focused his main attention on improving MTA finances, including new operating subsidies and a massive infusion of capital funds. By the end of 1981 the MTA headquarters team assembled by Chairman Ravitch had succeeded in establishing an important new transportation springboard for the resuscitation of the region's transit system. Making use of his understanding of the political realities of New York City and of the state capital in Albany, and indulging in unusual candor about the magnitude of the problems facing the MTA, Ravitch was able to convince the legislature to declare a "transit emergency" and to assemble a financing package that would assure adequate funding for the first Five Year Plan for the rebuilding of the basic transit infrastructure. In 1980, 1981, and 1982 Ravitch was also able to secure major increases in operating subsidies from a set of dedicated taxes approved by the state legislature.

Emerging from Crisis Management

The MTA Capital Program for 1982–1986 was meant as a first step in the restoration of the physical plant and rolling stock of the region's transit system. Completely rebuilding the subway, bus, and commuter rail networks; restoring the system to a "state of good repair"; and providing for the normal replacement of equipment and facilities would cost more than $1.5 billion a year until 2006. From this point on, an additional investment of $1 billion a year would be required for normal replacement of components of the physical plant and rolling stock (all in constant dollars).

Expenses of daily operation of the MTA system were projected to increase at a faster rate than general inflation, and operating subsidies were expected to remain stable or to decline. Management practices at all of the operating agencies needed to be overhauled, particularly at the Transit Authority and at the newly restructured Metro-North Commuter Railroad, which in 1983 took over from Conrail the direct operation of commuter rail service from the five counties north of New York City, Connecticut, and the Bronx to Grand Central Terminal.

The arrival of a new management team at the MTA and the TA under Robert Kiley in late 1983 and early 1984 meant that many of the programs and projects initiated under the previous leadership were carefully scrutinized. Often, the answers to their inquiries disturbed Kiley and his new managers. At the Transit Authority, funds were being wasted because of confused investment priorities and imprudent administration. Service on the TA subway system was still subject to serious delays because of the so-called "red tag crisis," as trains slowed to a crawl in areas where deteriorated track imposed severe speed restrictions. Meanwhile, day-to-day maintenance of the TA system was not being properly managed, which contributed to the poor service.

Management reform became the first priority. Work on portions of the capital program was suspended at the Transit Authority while the program was reviewed and management changes were initiated. With the support of the Mayor of New York City and of the governor, the MTA was able to win new management powers and to negotiate changes in restrictive work rules. Hiring of a new group of 1,200 non-civil-service managers was authorized. For calendar year 1985, the TA established dozens of targets for improvements in service and increased efficiency and effectiveness and was able to achieve most of its key goals.

The commuter railroads, the problems of which had been less severe, made further progress. Management reorganization at the LIRR brought the $1.1 billion capital plan under better control. The Metro-North Commuter Railroad, which weathered a 6-week strike just after it was formed in 1983, is being transformed into a modern commuter railroad; it has experienced ridership increases and substantial gains in service quality.

A Changing Region

Improvements in the management structure at MTA and the operating agencies began during a time of economic stability in the New York region. Clearly, the New York region has been on an upswing since the late 1970s. Economic growth in the region resulted in a wave of new development, both commercial and residential. Office and retail developers were aware of the problems of congestion and the limitations on the regional transportation system and were willing to undertake improvements to the transportation system in return for bonuses such as greater density or additional floor area. New York City has incorporated mandatory transit-related improvements such as relocation of subway entrances within the building line into its land use regulations, especially for highly congested Manhattan zones. The city's Uniform Land Use Review Procedure (ULURP), by publicly evaluating project environmental impact statement findings, has required developers to provide mitigation measures to offset the impact of high-density development projects, and MTA has been increasingly successful in securing transit-related improvements through negotiations with developers.

From 1977 on, the Manhattan CBD experienced steady increases in employment; by the early 1980s job growth had also resumed in the outer boroughs of New York. Meanwhile, the number of jobs in the New York suburbs continued to grow. This increasing employment, together with population growth and slower inflation, led to increased optimism about the future of New York City. The fiscal crisis appeared to be behind the city; the success of the Emergency Financial Control Board in stabilizing the fiscal condition of the city, and the more reliable stream of revenue from the city's tax base, led to the reentry of New York City into the financial markets for both short- and long-term borrowing.

Although the city was in better shape, some pointed out that, in many ways, it was not the same city. For one thing, the types of jobs available in Manhattan were changing. The fastest growing sector of the economy was the so-called FIRE (finance, insurance, and real estate) sector. Manufacturing jobs, not only in New York City but in the suburban portions of the region, had declined sharply. Service jobs were coming to dominate the economic picture, and the shifting patterns of employment—not only types of jobs but locations of jobs and the workers' residences—resulted in shifts in transportation demand.

During the first year of the SPI, several emerging regional trends were examined in detail to determine their potential impacts on the region's transit needs. Some of the social and economic problems facing the MTA region have to do with the city-versus-suburb dynamic. Between 1985 and 2000, the region's population is expected to increase by 325,100 or 2.5 percent. Most of this growth will take place in the suburbs; New York City's population will be stable or increase slightly, while the suburban counties and Connecticut will grow by 5 percent overall. Although Manhattan is tremendously influential, with the greatest concentration of jobs in the region, suburban residential and employment prospects indicate that relatively more economic muscle could be exerted outside the CBD. Between 1985 and 2000 the region will gain 485,100 jobs, an increase of 7.5 percent. Manhattan CBD employment will remain the most important sector, with the largest absolute increase, but suburban growth rates will be higher. The continued growth of employment in the suburbs, however, is accelerating traffic congestion in many locations, resulting in forecasts of suburban gridlock for some sectors.

In addition, an imbalance in the labor market is emerging as low-level service jobs go begging in the suburbs and inner city unemployment remains distressingly high. This situation is expected to continue for the remainder of the century. On the other hand, a disproportionate share of Manhattan CBD jobs has been taken by suburban commuters in recent years, leading to increased pressure on commuter rail and express bus services throughout the region. It is as yet unclear whether this trend will continue in the future. The increase in the employed labor force between 1985 and 2000 is expected to approach 6 percent with an almost 4 percent increase in New York City and a 10.6 percent increase in the counties to the north of the city. This may result in a net increase in commuter trips to Manhattan from these areas. On Long Island, however, the labor force is not expected to grow as fast as the number of jobs, resulting in the possible development of a significant "reverse commuting" pattern as New York City residents travel to jobs in Nassau and Suffolk counties. The net impact of possible changes in commuter flows has not yet been determined.

Meanwhile, the search for competitive advantage for the New York metropolitan area led to a realization of the critical importance of the region's transportation infrastructure. Although many categories of public investment in capital facilities were shortchanged during the fiscal crisis, the subway system had been suffering from disinvestment for decades, in part as a result of the artificially low transit fare. Restoration of the transportation infrastructure, especially the rapid transit system serving the Manhattan CBD, was perceived as a critical element in efforts to support economic expansion.

The national debate on the "Infrastructure Crisis" was the subtext for the search for solutions in the New York region. The Executive Director of the Port Authority of New York and New Jersey had already proposed the establishment of a "regional infrastructure bank," which would underwrite the long-term cost of improvements that could support economic development. In both New Jersey and New York, voters approved bond issues that established transportation improvement trust funds. At the federal level, a 5-cent per gallon increase in the federal gas tax included a 1-cent set-aside for mass transit projects.

At the same time, the traffic problem in the region presents a real opportunity for the MTA. Increasing attention is being paid to the escalating levels of congestion on the regional highway and street network. Increased automobile usage is slowly strangling traffic; transit is perceived by many influential observers as the only possible solution to the looming prospect of regional gridlock. In the near future, additional restrictions on the use of the automobile in Manhattan may help to boost transit ridership.

Opportunities also presented themselves in the form of the Westway debacle. After years of litigation and reams of impact analyses, the project intended to replace the inadequate and deteriorated West Side Highway was abandoned. The money allocated to this project was traded in through the Interstate transfer provision and funds were reprogrammed to provide for a more modest replacement highway and transit improvements.

Meanwhile, the efforts made by MTA and the operating agencies since the early 1980s to improve the transit system were starting to bear fruit. Visible improvements in service on the subway, bus, and commuter rail networks as a result of the first 5-year capital program—largely manifested by new subway cars, buses, and commuter rail cars—restored a certain sense that things could get better. Successful management reforms at various levels of the MTA and the operating agencies also helped to create a climate of cautious optimism.

It was also important to underscore the significant strengths that the MTA possessed. For the most part, the operating agencies had exceptional top-level management in place, as well as a skilled work force. The 5-year capital program that was nearing completion had helped to arrest the deterioration of service and represented an important first step in restoring the system to a state of good repair.

Even before the management changes sought by MTA Chairman Kiley were in place, it became clear that a new approach would be needed to help the MTA accomplish its purpose, which includes the development and implementation of a unified mass transportation policy for the MTA region. The MTA was facing the end of the first Five Year Capital Program, and many of the funding sources tapped for the first round of projects would be either depleted or unavailable. How could the MTA be certain the required $1.5 billion annual investment in the region's mass transit infrastructure would continue to be made?

Implementing a Strategic Planning Process at the MTA

Strategic planning had been seriously discussed for years at the MTA, but there had never been a comprehensive process, aside from the budget review process, that attempted to involve all of the MTA operating agencies in collective planning for the future. Strategic planning had entered the official debate when the state legislature in 1983 required the MTA to submit to the governor strategic operation plans for the transit authority and the commuter railroads.

In early 1985, Robert Kiley was searching for a mechanism that would galvanize the policy makers and the decision makers and that could evoke a broad consensus about the future shape of the mass transit network. The chosen vehicle was the strategic planning process, and the director of the MTA Planning Department was charged with the responsibility for making it work. A staff reorganization shifted lines of responsibility so that the Director of Planning reports directly to the MTA Chairman; the Director of Planning was to play a key role in coordinating the strategic planning program and making planning work at MTA. It was particularly important to mobilize the resources of the operating agencies and obtain the commitment of the top management of each agency.

The time frame for the SPI was set at 3 years, by the end of which all of the key issues were to have been identified and incorporated into the planning process. At the conclusion of the SPI process, strategic planning would be incorporated into the ongoing planning and budget process used at MTA and at the operating authorities.

Defining the Structure

The first order of business was to decide on the appropriate structure for the SPI process and to determine the roles of various institutional players. This involved the division of planning work within the MTA family of agencies, as well as coordination with outside planning departments and groups. In addition, because of the complex political structure of the MTA region, input from a broad cross section of elected officials and special interest groups had to be accommodated in order to provide for an open process.

Certain realities of the mission and mandate of the MTA affected the scope and level of detail that would characterize the SPI. The mission of the MTA was established very clearly in the authorizing legislation that created it: the continuance, further development, and improvement of commuter transportation and other related services, as well as the development and implementation of a unified mass transportation policy for the MTA region. Given this mission, how could the MTA marshal its resources and perform its mission efficiently, effectively, and with pride in its service to the public?

To initiate the strategic planning process, the MTA Planning Department developed a general outline of the issues that would have to be addressed, a rough timetable for implementation, and some guidelines on the basic economic and demographic projections that would be used in the course of the analysis. After this initial scoping effort, the Planning Department developed a draft work program for the first year of the SPI. This work program was discussed extensively with the MTA Board, agency heads, and planning directors in the spring of 1985. Out of these discussions came an agreement on the basic format of the SPI effort, as well as a detailed work program for year one.

Setting Assumptions

Early in the process of structuring the multiyear strategic planning process, MTA headquarters staff determined that a coordinated approach to the regional transportation problem was essential. The initial intent was to have a shared set of assumptions about the level of financial resources that would be available to the MTA in the foreseeable future. It was soon decided that a far more extensive framework of assumptions would be required to ensure that all of the constituent MTA agencies would be moving in the same direction as they executed the work programs that had been developed for the first year and beyond.

The list of assumptions started with the principle that the mission of MTA would remain the same under a wide range of external conditions—that is, that the delivery of service in the most efficient, cost-effective manner was the overriding goal of MTA. Next, it was assumed that the goal of the existing capital improvement program—to restore the system to a state of good repair and to maintain it at that level—was an unalterable commitment. It was also determined that service rationalization would be based on a principle of equity; that is, there would be well-defined guidelines for service delivery that would guide the operators in assessing how to allocate their resources.

A controversial element in the set of guiding assumptions was the statement that the level of available funding might be inadequate (in view of recent federal positions on mass transit budget levels) to restore the entire MTA system to a state of good repair. This articulation of a perceived threat opened the MTA to charges that service reduction was the overriding target of the exercise. Indeed, availability of capital funding was perceived by MTA staff as the principal element that threatened to undermine the goal of improved service. In a very real sense, the development of a set of less than optimal service configurations in response to possible funding constraints would serve to demonstrate how vital to the region the MTA transit network is.

Other components of the initial assumptions dealt with forecasts of population shifts, economic development, and inflation; and other elements introduced the issues of public sector–private sector cooperation and interagency service coordination. Still other assumptions stressed that deferred maintenance of facilities and equipment was not an acceptable strategy for dealing with potential operating or capital funding shortfalls. In addition, each agency was encouraged to investigate potential savings from the introduction of new technology or through increases in productivity stemming from management improvements and better utilization of labor.

Development of the first year assumptions both restricted the choices available to planners at the operating agencies and freed them from expending time and energy in developing detailed forecasts and projections of population and economic activity. The work program and the overall framework of assumptions were flexible to the extent that the planning groups at the individual operating agencies could modify them in consultation with the MTA Planning Department if there were a compelling reason. Planners would propose to modify fare policy, for instance, in order to encourage the use of certain services or connections or to accommodate passengers inconvenienced by service adjustments. Such proposals would have to include an analysis of the costs and benefits associated with the policy change.

Steering the Process

The institutional relationships among the different elements of the MTA—headquarters staff, operating authorities, and even interdepartmental relations at the different operating agencies—were not necessarily aligned so that the strategic planning program could proceed vigorously from the very start. Some delays could be anticipated while the operating agencies and the MTA sorted out the responsibility for the delivery of particular products related to the overall work program.

An important element in the eventual success of the SPI was the involvement of the MTA Board and the commitment of the Chairman and Chief Executive Officer of the MTA, Robert Kiley. Kiley had essentially delivered the mandate for the process by directing the MTA Planning Department to undertake the study. The involvement of the MTA Board was ensured through two mechanisms: the establishment of the Planning Committee of the MTA Board and a briefing held in April 1985 to discuss the SPI.

Several mechanisms were developed to keep completion of the first year work program on schedule. The work program was a relatively demanding enterprise because, in some respects, the individual agencies would be getting a "standing start" on many of the components of the agenda. To encourage timely submission of scheduled work products and to share the knowledge gained as each agency advanced along the learning curve, a series of "Planning Directors' Roundtables" was held. In addition to sessions devoted to progress reports on the products of the SPI work program, special meetings were scheduled for discussions of West Side corridor issues, privatization, and automatic fare collection.

When the initial outline for the first year work program had been prepared, a 1-day board briefing was held to discuss regional transportation issues and to familiarize the board and agency presidents with the SPI. Much of the discussion at the briefing centered on a series of assumptions that were prepared by the Planning Department to frame the discussion and to provide a common starting point for studying regional problems that faced the MTA. These assumptions are discussed in more detail in the next section. Although the assumptions were generally supported, there were significant changes incorporated in the assumptions as a result of the board's review.

By the time the board briefing was held, the planning departments of the operating agencies had had an opportunity to review the proposed work program for the first year of the SPI. After the briefing, each of the MTA agencies prepared a detailed work program for the first year, based on the instructions but modified according to the particular circumstances of each organization. These work programs were finished in June 1985. The various MTA agencies had different degrees of experience with long-range or strategic planning, and past management decisions had influenced the priorities and emphasis areas that affected relative positions on the strategic planning learning curve. Consequently, the agencies committed themselves to accomplishing different portions of the work program in year one.

Planning at MTA: Independent Studies and Coordination

Aside from the work programs generated by the operating agencies, the MTA Planning Department and headquarters staff retained responsibility for certain overall policy or service coordination issues and for analyses of particular transit corridors where a number of different governmental and private actors were involved. For example, although the MTA encouraged each operating agency to consider how elements of its service package could be improved through private-sector involvement, the MTA Planning Department staff initiated a policy analysis of the institutional, legal, labor-related, and other issues that would arise if the MTA wished to place greater reliance on private-sector transportation providers. The MTA also sponsored a corridor study to assess the desirability of making provisions for a transit right-of-way on the West Side of Manhattan. The state of New York was at that time preparing to draw up plans for a replacement for the abandoned Westway project, and a unique opportunity presented itself to provide for a future high-volume transit service, whether bus, light rail, or heavy rail, to serve the expanding West Side.

In addition to these policy studies, a methodology known as the Subway Service Utilization Model (SSUM) was developed by the MTA Planning Department during the first year of the program. The SSUM was designed to calculate and display the levels of volume and crowding on the New York City subway system. Although the methodology was not intended to be used as a basis for making specific service changes, it could indicate in a general way where some of the problems and opportunities for the system lie. A final report detailing the capabilities of the model was completed and published in September 1986.

Two major, independent planning efforts were also sponsored as part of the first year program. The Regional Plan Association, a nonprofit research organization with a solid background in regional transportation issues, was retained by the MTA to provide an independent assessment of the demand for transit in the New York area. Also retained by the MTA, the Urban Research Center of New York University, an academic research arm of a major metropolitan university, undertook an analysis of future levels of demand for service and strategies to provide service under limited funding. These independent efforts were intended to provide a perspective on regional problems that might not have been shared by the MTA or its subsidiaries, given the necessary orientation toward provision of existing services that characterized the operating agencies.

Community Participation

The MTA has an obligation to reach out to the public, but it is even more essential to find out what customers want in order to serve them better. The customers or clients of the MTA include not only passengers, but potential passengers (those who do not now use the services); indirect beneficiaries (those who do not themselves use the system but who benefit in some way because other people use it); and funding sources (elected officials and representatives of funding agencies at various levels of government).

The SPI has a substantial public participation component that will continue in the second and third years of the program. Key elected and appointed officials were informed of the SPI before its public announcement, and extensive briefings were held throughout the region to provide additional details and to solicit comments. Public forums on the SPI and strategic planning in general have been sponsored by civic and advisory groups, and presentations about the program have been made to interested groups. Press attention, including editorial comment, was considerable and generally positive.

As a supplement to the presentations to local elected officials and community representatives, a response form was distributed to community board members to solicit each area's top transit priorities. Key concerns raised by this procedure will be considered during the second and third years of the program. The areas of concern include transit's ability to enhance local economic development; transfer policy, including bus-to-subway transfers; reverse commuting; and achievement of cost savings through productivity and technological innovation.

Strategic Planning at the Operating Agencies

The TA was in the process of setting up a new organizational structure for planning, which would incorporate an Operations Planning Group within the TA's operational arm and a Strategic Planning Department in the finance and administration branch. Although the overall management of the TA's portion of the SPI would be administered by the Strategic Planning team, operations planning would be important in areas such as developing service guidelines and system reconfiguration analysis.

The two agencies were in a good position to carry out the strategic planning program with the commuter rail lines. The MNCR had already commenced work on some of the elements of the work plan, and service guidelines were already under development. The Metro-North planners also recognized that the development of a methodology for determining cost impacts of potential service changes would be extremely valuable. The LIRR had already considered long-range business planning important. Through the SPI, LIRR would seek to meet its corporate goals of improved service and financial performance.

The MSBA and the TBTA took somewhat different approaches to the SPI than did the TA and commuter railroads. MSBA, which provides bus service in Nassau County, was facing an increase in ridership and extreme constraints on available funding. Consequently, MSBA was dealing with a growth scenario in which service expansion might not keep pace with demand. TBTA, which operates toll bridges and tunnels within New York City, was also facing traffic increases and had to determine the optimal use of existing capacity. Because the TBTA generated a surplus that was applied to the operating and capital needs of the MTA's other services, the goal was to accommodate growth while improving efficiency and cost-effectiveness.

The first year of the SPI was quite revealing, not only in terms of the data collected and analyzed but in the improved understanding of the political, economic, and social realities that will influence transit policy development in the MTA region. An important component of this new insight was the discovery of a range of opportunities for improving the transit system and for potentially increasing ridership.

A generally shared perception emerged, which was strengthened by the results of the first year work plan. By and large, the New York City subway system was heavily used during the peak period (especially the morning peak period) for the journey to work. This implied that changes to the subway system could be made only at the margins of the system and that the basic configuration of the rapid transit system would probably remain unchanged. No line eliminations were likely. However, the surface transportation system was almost an unknown quantity. Few data were available on ridership, load patterns, and crowding. A key aspect of the first year work program was the development of service guidelines for the TA bus network, as well as the implementation of a comprehensive data collection effort.

Another factor that became apparent during the first year of the SPI was the difficulty of arriving at a consensus regarding changes to transit service. The not-in-my-backyard syndrome means that, even as people clamor for improved service, neighborhood groups campaign relentlessly against the location of new facilities or the addition of more traffic in existing corridors. This phenomenon is not limited to transit, as observers of waste disposal issues know. However, it is difficult to reconcile demands for more bus service with intense opposition to the construction of a new bus depot, or the clamor for reduced crowding on heavily used subway lines with opposition to the recycling of underutilized rail rights-of-way.

The same problem becomes even more acute if service reduction is considered. Tremendous pressure is brought to bear through the political process and the activities of transit advocacy groups if any mention is made of service adjustments or reconfiguration because the immediate conclusion is that service adjustment equals service reduction. This problem is complicated because the recent history of service changes almost confirms this perception. The TA, for instance, has in recent years adopted a de facto "zero-sum game" policy toward service adjustments. In other words, any increase in service anywhere in the system must be balanced with a commensurate reduction in service elsewhere.

In an effort to look at service in a comprehensive fashion, the MTA had asked the operating agencies to examine segments of the system that are relatively underutilized to determine whether the curtailment of service or the elimination of facilities would result in significant capital savings. This analysis was suggested because many portions of the system are in need of extensive rehabilitation, and there is no assurance that sufficient funding will be available to fully reconstruct the system in its present configuration. Prudence dictates that the TA invest its resources in places where they will serve the greatest number of riders and that priorities be set on a cost-benefit basis when decisions about service adjustments are made. Even so, considerations of equity cannot be discarded; in some instances allowances would have to be made in response to socioeconomic conditions that contribute to transit dependency.

Even to suggest alterations to line segments, however, opened the MTA and the TA to accusations of a hidden agenda and resulted in the revelation of a so-called "secret plan" to cut service when a draft of a TA report was leaked. In fact, a preliminary analysis of the line segments under study indicated that no significant cost savings would be realized by closing the lines, particularly because some facilities such as subway yards and maintenance facilities are located at the ends of subway lines. These facilities would be expensive to relocate. In addition, substitute bus service would probably be more expensive to provide on a unit cost basis.

The SPI was conceived as a multiyear effort from the start. Experience with the strategic planning process elsewhere indicated that it takes some time to establish momentum and to accumulate the kind of information required to make long-term decisions about corporate direction. The question of a data base was particularly significant because much of the basic information on how the subway and bus system is used was unavailable when the program began.

The lessons from year one will be incorporated into the Second and Third Year Work Programs of the SPI. One key piece was the use of the data that have been gathered to assess policy alternatives. The completion of the data collection and technical analysis incorporated in the first year program was extremely important, especially because the work plan had been developed to provide vitally needed tools for rationalizing and improving service.

Among the most useful products of the first year was the creation of two sets of service guidelines for local bus service provided by the TA. Previously, decisions about levels of service were made on an arbitrary basis, and no formal standards were in place. Schedules and route structure decisions were essentially made according to the professional judgment of operations personnel. Under the new system of service guidelines, criteria were established that consider population density and measures such as automobile ownership in determining route structure and that take into account passenger loads and cost-effectiveness in adjusting frequency and span of service.

Part of the task of the SPI is to reconcile the different mobility needs of various parts of the region. The city of New York is sometimes described as "two cities": the Manhattan CBD as a Gold Coast, with vast concentrations of wealth and jobs, and the rest of the city, with pockets of poverty and far less weight in the decision-making process. Although the needs of the "two cities" are different, each of the outer boroughs has unique strengths in terms of the industries, transportation infrastructure assets, and community facilities that they possess. Within the city, a key issue is the level of service required by the residents who work at job sites outside the Manhattan CBD. This issue is critical because many of these residents do not have access to automobiles. Almost 60 percent of New York City households had no automobile in 1980, and fewer than 20 percent of Manhattan households owned an automobile. In contrast, only about 10 percent of households in the New York suburbs were without an automobile. In the first year of the SPI, the Office of Strategic Planning at the TA studied the characteristics of major employment centers outside of the Manhattan CBD. A report published in January of 1986 entitled Non-Manhattan Employment Center Work Trip Patterns surveyed journey-to-work data from the 1980 census as part of the initial data collection and analysis effort.

It also became apparent that it is essential for the MTA to establish a set of shared assumptions, a common view of where the region is heading and of the role of the MTA in making the region more competitive. Businesses were increasingly considering relocation outside New York City or outside the region, and the inadequacy of transit to serve their needs was frequently cited as a factor in relocation decisions. A coordinated transportation system, integrating bus, rapid transit, and commuter rail service, could not be achieved unless planners and decision makers agreed on some fundamental issues, such as the anticipated economic prospects for the region, the likely corridors of growth, and the proper role of the different modes in areas where service overlaps.

An important component of the first year program had to do with improving the coordination of one MTA agency with another, as well as the relations of MTA with other transportation agencies in the region and with local, state, and federal funding sources. The different operating agencies under the MTA umbrella historically have failed to work as a team, and the MTA had long been criticized by outside agencies for not putting its house in order and for failing to cooperate with other providers of transportation services such as the private bus lines in New York City or the Port Authority of New York and New Jersey.

FUTURE OF STRATEGIC PLANNING AND MANAGEMENT AT THE MTA

A great deal of progress was made in the first year of the SPI. Each of the operating agencies made strides in improving data collection capabilities, in assessing the likely consequences of reduced capital funding availability, and in internalizing the concept of strategic planning. The decision makers at MTA and the operating subsidiaries achieved a greater understanding of the implications of growth and change in the region. The MTA also made progress in convincing outside critics that a serious examination of the strengths and weaknesses of the organization, as well as the threats and opportunities facing the MTA, was under way. The next challenge was to maintain momentum and to build on the successes of the first year.

Activities that will continue include data collection efforts, completion of development of service guidelines, and completion of the system reconfiguration analysis that will respond to the different potential levels of capital funding available. Additional work will be done on assessing the costs and benefits of changes in service, developing methodologies for calculating operating costs and ridership impacts of service adjustments, and studying demographic and economic changes throughout the region.

The work plan for the second year of the SPI builds on the results of the first year, but it also extends to the development of a methodology for deciding how to reconfigure the system. The proposed methodology, a corridor planning approach, would address the issue of service coordination, the problem of how to utilize available capacity, and the response to changing levels of ridership. The corridor studies would serve as a screening process for identifying problem areas, developing a variety of service improvement proposals, and getting an order of magnitude estimate of the potential for service improvements or cost savings.

When the corridors have been screened, the highest priority corridors can be studied in depth. Detailed alternative service plans will be developed by the operating agencies. Among the criteria used to select the best alternative would be increased ridership, improved service to existing riders, reductions in capital or operating expenditures, and contributions to other societal goals such as reduced air pollution and less congestion.

The corridor studies themselves are intended to encourage innovation in the use of unconventional solutions to problems. Fare policy, demand management through staggered work hours, mode substitution, and privatization are among the potential methods available to attract riders, reduce peak-period crowding, speed service, or reduce costs. The regional transit system needs to function more as a network, and these methods and practices will lead to a better-integrated transportation system.

The intent of this open process is to permit the re-evaluation of planning assumptions. One of the first steps in preparing a second year work program was the reconsideration of the assumptions adopted for the first year of the program. The experience of the first year may have invalidated some of the initial theories, and new information or newly emerging trends may have resulted in new hypotheses. Even fundamental issues such as ridership goals were open to question. For example, the issue of whether it should be an MTA goal to increase ridership in all transit markets—that is, in all parts of the MTA region and at all times of the day—has been raised. The answer depends on the availability of other service in those areas, the cost of accommodating additional passengers, and other factors.

The SPI is a flexible process, and the work program will be expanded as needed to address problems that emerge in the course of the study. Each agency must constantly re-examine the basic questions: "What business am I in? How am I doing? How do I know? How can I be more effective?" The strategic planning process should help the managers of the system answer these questions. The answers will affect what kinds of questions have to be asked in the next cycle.

Even after the end of the SPI, strategic planning should be part of the overall decision-making process. One way to integrate strategic planning is to include it in the budget cycle. Early in the SPI program, a directive by NYCTA President David Gunn requiring that all budget submissions be justified in terms of SPI goals indicated that the SPI process was being internalized at the TA, rather than remaining an extraneous, pro forma exercise. The products of each year's strategic planning effort should be the identification of key opportunities and areas of emphasis. These priority issues will be emphasized as the agency prepares its budget for the following fiscal year.

At the end of the formal SPI process, a series of reports will have been issued and a number of studies will have been completed. The process of optimizing the region's mass transit resources will have begun. The real product of the process, however, should be the implementation of an ongoing planning and budgeting system that incorporates strategic management from the very start. In order to do their jobs better, all of the MTA's staff will be "thinking strategically" as they respond to change.

Publication of this paper sponsored by Committee on Strategic Management.

Initiating the Strategic Planning Process at NJ Transit

CHRISTINE BISHOP-EDKINS AND CYNTHIA NETHERCUT

The adoption of a strategic planning process is becoming more important to public transportation agencies as the industry faces declining subsidies and increased competition from other carriers as a result of deregulation. Valuable as strategic planning may be, processes and structures developed by and for the private sector are unsatisfactory for the public transit industry. During the past year and a half, New Jersey Transit Corporation (NJ Transit) has been adapting these processes and structures for its own strategic planning efforts. In this paper is described NJ Transit's strategic planning process, which involves critically assessing NJ Transit's opportunities and threats in a market and its performance relative to competitors. These assessments are used to position NJ Transit's services in a matrix that recommends strategic roles and actions. Strategies and resource allocation decisions are then based on the location of services in the matrix.

American corporations were introduced to formal strategic planning in the mid-1950s. Since that time, strategic planning has become so widespread that managers of most large corporations around the world practice it in some form. According to a recent survey, strategic planning ranked as the most important responsibility of 62 percent of the chief executives of the 550 largest industrial, banking, diversified financial, life insurance, retailing, transportation, and utility companies identified by *Fortune* (*1*, p. 36). Strategic planning, the systematic identification of future opportunities and threats and attendant strategies, appears firmly entrenched in business management.

In contrast, only a few state and local transportation agencies use strategic planning systems to help allocate resources among their different services. The Pennsylvania Department of Transportation, for example, has restructured its operation using strategic planning to initiate productive activities. The Port Authority of New York and New Jersey is using strategic planning to reconceptualize and manage its transportation businesses so that they support economic development objectives in the New York metropolitan region (*2*, p. 20).

INITIATING STRATEGIC PLANNING AT NJ TRANSIT

NJ Transit is a statewide public transportation agency created by an act of the New Jersey Legislature in 1979 to manage and improve bus and rail passenger services throughout the state. During the first few years of its operation, the agency was primarily concerned with improving a transit system characterized by declining ridership, attributable in part to deteriorated services, equipment, and facilities. NJ Transit delivers bus and rail services under the auspices of three operating subsidiaries created between 1980 and 1984. Under its first subsidiary, NJ Transit Bus Operations, Inc., bus service is provided to 20 of the state's 21 counties on a variety of routes ranging from local urban routes to long-distance commuter runs to Newark, New York, and Philadelphia. Approximately 430,000 daily passenger trips were taken on NJ Transit buses in 1985. NJ Transit's second subsidiary, NJ Transit Rail Operations, Inc., provides commuter rail service in New Jersey. In 1985 NJ Transit served 150,000 daily passenger trips on its nine railroad lines spanning 12 counties. A third operating subsidiary, NJ Transit Mercer, Inc., operates the former Mercer Metro bus system in Trenton.

In January 1985, NJ Transit formally marked its transition to an agency planning for the future by hiring AT&T's Organization Effectiveness Group to teach their strategic planning process. At a 3-day conference facilitated by AT&T, managers analyzed environmental trends, evaluated the strengths and weaknesses of NJ Transit's services compared with those of its competitors, and formulated strategies to take advantage of external opportunities and internal strengths (*3*). Major products of the conference included initial mission statements and action plans for bus and rail operations.

NJ Transit was unable to complete an analysis of its strengths and weaknesses at the conference, in part because AT&T's process was inappropriate for analyzing a public-sector transportation agency. Nevertheless, top management believed that an in-depth analysis of bus and rail services should be performed. To this end, NJ Transit's Office of Strategic Planning adapted various strategic planning processes to meet the particular needs of transit.

USING THE MATRIX AS A DECISION-MAKING TOOL

Portfolio Evaluation

Since the late 1960s, several large corporations, among them General Electric, Mead, and Olin, have been using a strategic planning device called portfolio evaluation to help them make investment choices among different product lines, companies, or divisions (*4*, p. 3). Portfolio evaluation identifies the contribution of the corporation's business units (product lines, companies, or divisions) to overall performance and clarifies their roles. Management can then decide which business units should be used to generate cash and which should receive investment funds (*4*, p. 3).

C. Bishop-Edkins, Marketing Department, New Jersey Transit Corporation, Inc., McCarter Highway and Market Street, Newark, N.J. 07101. C. Nethercut, Finance Department, Regional Transportation Authority, One North Dearborn Street, Suite 1100, Chicago, Ill. 60602.

Investment choices among the individual business units have typically been based on the unit's projected profitability and market share and depicted on a growth/share matrix. Strategic roles for the units accompany the matrix and suggest specific management actions. Many variations of this matrix have resulted, including approaches by Arthur Little, Inc., the General Electric Company, and AT&T (*4*, p. 3).

MacMillan's Matrix

Ian C. MacMillan of the Wharton School of Business, the University of Pennsylvania, recently developed a matrix to guide resource allocation decisions in not-for-profit agencies. MacMillan asserts that it is much more difficult for not-for-profit organizations than for private industry to decide how to allocate extremely limited resources because service agencies must choose among a portfolio of needy programs. Whereas discontinuing a product may cause only minor inconveniences for former customers, eliminating or trimming necessary social services can cause human suffering (*5*). Contrary to private-sector–oriented models, MacMillan's matrix incorporates the complex allocation alternatives faced by public service agencies.

MacMillan's matrix is used to analyze all current and potential programs on the basis of program attractiveness, competitive position, and alternative coverage. These three major dimensions determine the location of an individual program in the matrix, shown in Figure 1, and the role the program plays in the overall portfolio of social service activities.

The basic assumptions of the matrix make MacMillan's approach appropriate for public transportation agencies. For nonprofit agencies to survive, they must be willing and able to compete for limited resources with other agencies. Because resources are limited, agencies should not directly duplicate others' services thereby wasting resources and creating inefficiencies. This situation requires sacrificing some duplicative, low-quality programs to provide quality service to more focused markets (*5*).

Given these assumptions, the ideal portfolio or mix of programs is one in which an agency only serves markets in which its competitive position is strong. An agency builds this ideal portfolio by easing weaker competitors out of markets it can serve better and by conceding its weaker programs to stronger competitors. A social service agency is required to serve both attractive and unattractive markets as long as it is the superior provider. The agency uses attractive programs to support less attractive programs that have no or few alternative providers.

NJ Transit's Matrix

NJ Transit adopted salient features from both MacMillan's and AT&T's approaches. MacMillan's matrix was modified to create NJ Transit's matrix as it appears in Figure 2. The interpretation of the matrix cells is given in Table 1. NJ Transit then adapted both MacMillan's and AT&T's processes for evaluating and placing services in the matrix. The agency's various bus and rail services were arrayed on this matrix according to three dimensions: market attractiveness, competitive position, and alternative coverage. Market attractiveness, as defined by this matrix, is the degree to which services are able to cover costs through fares or subsidies. Competitive position is the degree to which NJ Transit is superior to its competitors. Alternative coverage is the extent to which alternative transportation agencies could serve riders if NJ Transit ceased operating.

As the matrix suggests, a public transportation agency such as NJ Transit operates like a private operator in some markets and provides a necessary (unprofitable) public service in others. Like a private operator, in attractive markets with no or few providers, NJ Transit is free to expand its services unhindered by a competitor. But NJ Transit must compete aggressively for the whole market or for selective submarkets (routes) to maintain its services if it competes with another provider. On the other hand, as a public agency, it must provide service in unattractive but necessary markets where riders have no other travel alternatives.

MATRIX INPUTS

Market Segmentation

As a first step, NJ Transit divided its market into appropriate market segments. A market segment was defined as a group of

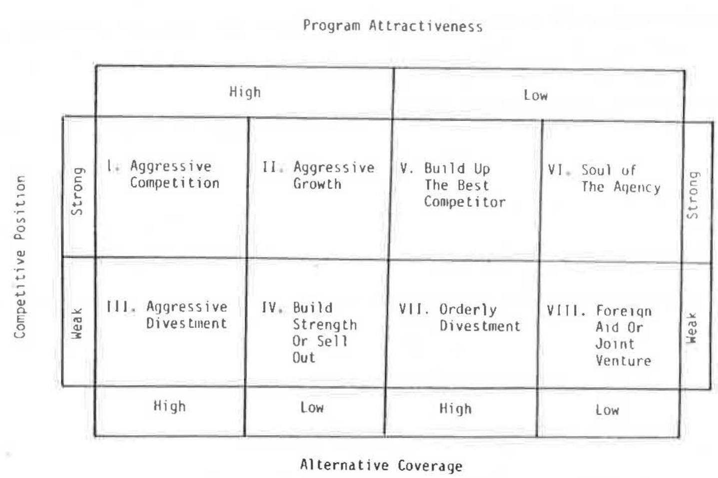


FIGURE 1 MacMillan's matrix.

TABLE 1 MATRIX CELL INTERPRETATION

Cell No.	Name	Definition	Primary Features	Strategic Imperatives
I	Aggressive competition	Strong competitive position Attractive market to serve Many alternative providers	Many transportation providers are competing for riders in an area in which NJ Transit has clear superiority	Identify key competitive variables (such as speed) and build these capabilities to capture the market; use these services to provide funds for growth in other markets
II	Aggressive growth	Strong competitive position Attractive market to serve Few alternative providers	The market is wide open to NJ Transit	Expand services rapidly and build competitive capabilities to ward off future competition Provides a reason for future existence
III	Contract out or exit market	Weak competitive position Attractive market to serve Many alternative providers	There are many providing services similar or superior to NJ Transit	Using the least amount of resources possible, NJ Transit should ensure the competition provides high-quality service
IV	Build strength or contract out	Weak competitive position Attractive market to serve Few alternative providers	Although these services have been recently initiated to fulfill a growing need, NJ Transit lacks the resources and skills to be competitive even in the absence of competing providers	If necessary resources are unavailable to respond effectively, encourage others to assume the service through contracting out; if resources are available, the service may be moved into Cell II or other position
V	Aggressive service maintenance	Strong competitive position Moderately attractive market Many alternative providers	Many competitors are providing services but to different degrees	Aggressively maintain all current services in the market to preserve strong competitive position
VI	Selective growth	Strong competitive position Moderately attractive market Few alternative providers	Opportunities may exist to expand or develop services because few providers exist in an attractive market	Expand services if the market can absorb them and NJ Transit can provide them economically
VII	Prove viability	Weak competitive position Moderately attractive market Many alternative providers	NJ Transit is a poor competitor in a moderately attractive market that is served by many others	Because others serve this market better, NJ Transit must justify its presence or contract out the service and exit market
VIII	Restructure service or contract out	Weak competitive position Moderately attractive mark Few alternative providers	Although there are few competitors in a moderately attractive market, NJ Transit is still in a weak position that could be the result of inefficiencies and misallocation of resources	Help other modes and operators provide service so NJ Transit can exit If services cannot be replaced, decide to maintain, reduce, or end service • Is the service necessary? • Does the service make financial sense? • Could better service be provided if NJ Transit reduced and focused service? • Are there alternative ways of providing the service?
IX	Selective service maintenance	Strong competitive position Unattractive market to serve Many alternative providers	Nonproductive competition occurs between providers who vie for a market share	NJ Transit should only maintain services that cannot be provided as well by another operator Leave remaining services to other operators
X	Soul of agency	Strong competitive position Unattractive market to serve Few alternative providers	Riders have no other services to depend on and, because the market is unattractive, it is unlikely another provider would appear	Pursue creative ways to provide these services Find ways to use other services to support those that fall into this cell
XI	Orderly divest or contract out	Weak competitive position Unattractive market to serve Many alternative providers	Market is unattractive and the services of many other providers are superior to those of NJ Transit	Concede these services to another provider Ensure smooth transition of riders from present services to competitors so that there is minimum disruption to riders
XII	Joint ventures	Weak competitive position Unattractive market to serve Few alternative providers	NJ Transit may be required to provide service for political or social reasons	Transfer riders to alternative providers if possible and give support to these services

Competitive Position	Market Attractiveness: High		Medium		Low		
Strong	I. Aggressive Competition	II. Aggressive Growth	V. Aggressive Service Maintenance	VI. Selective Growth	IX. Selective Service Maintenance	X. Soul of The Agency	Strong
Weak	III. Contract Out Or Exit Market	IV. Build Strength Or Contract Out	VII. Prove Viability	VIII. Re-Structure Service Or Contract Out	XI. Orderly Divest Or Contract Out	XII. Joint Ventures	Weak
Alternative Coverage	High	Low	High	Low	High	Low	

FIGURE 2 NJ Transit's matrix.

riders who have similar travel behavior and system use. A list of possible criteria for segmenting riders into markets included: geography, trip purpose, direction of travel, time of day, destination, and type of service. NJ Transit found it easiest to evaluate its services by segmenting its markets according to geographic criteria because data are generally collected by rail line or by bus route groups.

Market Attractiveness

After appropriate market segments were chosen, the market attractiveness of each was evaluated. A market segment is attractive from transit's perspective if it generates enough revenue to cover costs. Revenue for public transportation can be obtained through fares and, if there is political support, from state and federal subsidies. Thus, both economic and political criteria are used to determine market attractiveness. The criteria used to judge market attractiveness and their definitions appear in Figure 3.

To show how the market attractiveness of a market segment is determined, the evaluation of one of NJ Transit's bus segments, the Short Distance PABT routes, is given in Table 2. These routes run between densely populated New Jersey cities in Essex, Hudson, and Bergen counties and the Port Authority Bus Terminal (PABT) in nearby Manhattan.

After the segment had been rated according to each of the criteria, five or fewer of the most important determinants of market attractiveness for the segment were chosen. Overall market attractiveness (high, medium, or low) was based on an average of the most important criteria. Because the Short Distance PABT segment received a high rating for four of the five most important criteria, the overall market attractiveness of the segment was rated high.

Competitive Position

Next, the ability of each segment to fulfill rider needs and wants was compared with that of other modes such as automobile, rail, private buses, and vanpools. The definitions of the travel attributes important to users of all forms of transportation are presented in Figure 4.

In the Short Distance PABT market, NJ Transit competes with private cars and fixed-route vans. As the data in Table 3 indicate, NJ Transit's bus services were first analyzed to determine if they met the criteria to a high (H), medium (M), or low (L) degree. This same analysis was performed for each competitor. An H, M, or L was written in the first column to indicate the absolute degree to which both NJ Transit and the competitor met customer needs and wants. For example, NJ Transit's buses and private vans are considered moderately reliable whereas cars are believed to provide a high degree of reliability.

The numbers in the second column of each group indicate the relative competitiveness of NJ Transit's services and those of each transportation provider in a market. A 1 was assigned to the superior competitor, and the provider without the competitive advantage was given a 0. Both were assigned 0 if

Economic Factors
Revenue/cost ratio: A segment with a high revenue/cost ratio is attractive.
Concentration of riders: A segment with highly concentrated riders is attractive.
Absence of competitors: A segment with no competitors is attractive.
Proximity to NJ Transit facilities and infrastructure: A segment close to current facilities is attractive.
Condition of facilities: A segment with good facilities is attractive.
Land use: A segment supporting efficient land uses is attractive.
Political Factors
Rider influence: A segment is attractive or unattractive depending on the effectiveness of rider support or criticism.
Market share: A segment with a large market share is attractive.
Number of riders: A segment with many riders is attractive.
Appeal to stakeholders: A segment with appeal to those who either affect or are affected by transit operations or policies is attractive.
Mobility for transit dependent, seniors, and disabled: A segment providing mobility to these persons is attractive.
Quality of service: High service quality is attractive.
Economic and Political Factors: A segment experiencing growth in ridership is attractive.

FIGURE 3 Definition of key market attractiveness terms.

TABLE 2 MARKET ATTRACTIVENESS

	Short Distance PABT		
	High	Medium	Low
Economic criteria			
Revenue/cost ratio[a]	X		
Concentration of riders[a]	X		
Absence of competitors[a]	X		
Proximity to NJ Transit facilities, infrastructure	X		
Condition of facilities			X
Political criteria			
Rider influence	X		
NJ Transit's market share[a]	X		
Number of riders	X		
Appeal to stakeholders		X	
Mobility for transit dependent, elderly, and handicapped	X		
Quality of service[a]		X	
Economic and political criteria: growth rate of riders[a]	X		
Overall attractiveness	X		

[a]Criteria considered the most important determinants of market attractiveness.

Comfort: Physical conditions in the vehicle
Convenience: Number of transfers and ease of transfer (or ease of movement between modes)
Reliability: Arrival at destination on schedule
Safety: Perception of accidental injury or death
Security: Perception of incidence of crime
Cost: Entire cost of travel including parking and all transit fares or all car operating costs
Accessibility: Ease of traveling to and from major mode
Proximity to destination: Distance from major mode drop-off point to ultimate destination
Travel time: Door-to-door travel time
Frequency: Degree of flexibility in departure and arrival times
Lack of stress: Travel situations have various degrees of stress

FIGURE 4 Definition of key terms related to rider needs and wants.

neither service was superior to the other. Because the automobile is more reliable than NJ Transit's buses, a 1 appears in the second column under automobile and a 0 in the Short Distance PABT column. Vans and buses are equally reliable so both have 0 in the second column.

Because anticipated service modifications can change present assessments, the third column in each group was used to predict future competitive advantage. A 1 was assigned to the competitor who was expected to remain or become superior in the future because its performance probably could not be imitated by the other. A 0 was given to the provider without the competitive advantage or to both providers if neither was expected to capture the competitive advantage. In the latter case, 0 in the third column for both competitors indicates that their services were expected to remain or become equal to each other. The 0s in the third column of ratings for Short Distance PABT buses and automobiles indicate that in the future both will be equally reliable. Buses are expected to become as reliable as cars when a second express bus lane or other remedy is implemented to reduce bus delays in the Lincoln Tunnel. When buses were compared with vans, buses were assigned a 1 and vans a 0 in the third column because reduced bus delays will make buses more reliable than vans in the future.

In summary, the analysis in Table 3 shows that in the Short Distance PABT market the automobile is superior or equal to NJ Transit's buses in all areas except safety, cost, and lack of stress. And, with the exception of travel time, the automobile is expected to be superior or equal in those areas in the future. NJ Transit buses are much more competitive with vans, however. Buses are equal or superior to vans on all counts, and their superiority is expected to increase in the future.

Alternative Coverage

As was mentioned earlier, the alternative coverage dimension indicates the extent to which other transportation providers could serve riders if NJ Transit ceased operating. On the basis of its knowledge of competitors in the state, NJ Transit judged the probability of alternative coverage as high (H) or low (L) for each market segment. Vans and cars could probably replace NJ Transit's Short Distance PABT bus service, so alternative coverage for this segment was rated as high.

Placing the Services in the Matrix

As shown in Figure 5, the Short Distance PABT segment was located in a high market attractiveness cell because of its rating. It was also positioned to reflect high alternative coverage. Locating the segments according to strong or weak competitive position, however, was more complex. Although the analysis of rider needs and wants (Table 3) indicates areas of competitive advantage, it does not determine overall competitiveness. One approach is to base competitiveness on the services' market share because it accurately reflects which services people believe are superior.

For trans-Hudson trips, NJ Transit used data that describe the modes used by people traveling from specific geographic corridors in New Jersey to Manhattan to determine the bus and rail market shares of all trans-Hudson trips. Trans-Hudson services were considered strong competitors if they were

1. Dominant in a market: NJ Transit captures 50 percent or more of all trips to Manhattan from a specific corridor.
2. Dominant in a submarket: NJ Transit captures more than 50 percent of all trips to either midtown or downtown Manhattan from a specific corridor.
3. Equal to other modes: NJ Transit and its competitor or competitors are the strongest in the market and capture equal shares of all trips to Manhattan from a corridor.

Because 58 percent of trans-Hudson commuters from the area served by NJ Transit's Short Distance PABT bus service use buses, this mode is dominant in the market and thus a strong competitor.

Judging the competitiveness of local bus service compared with other modes was more difficult. According to the 1980 census, an average of 5 percent of all New Jersey intracounty work trips are taken by bus. Using intracounty bus work trips as an estimation of local bus patronage, NJ Transit decided that bus is a strong competitor if it captures more than 5 percent of all work trips.

When the market segments had been analyzed and rated for market attractiveness, competitive position, and alternative

TABLE 3 COMPETITIVE POSITION

Rider Needs and Wants	NJ Transit (Short Distance PABT)	Competitor (automobile)	NJ Transit (Short Distance PABT)	Competitor (van)
Comfort	M 0 0	H 1 1	M 0 0	M 0 0
Convenience	M 0 0	M+ 1 1	M 0 0	M 0 0
Reliability[a]	M 0 0	H 1 0	M 0 1	M 0 0
Safety	H 1 1	M 0 0	H 1 1	M+ 0 0
Security	M 0 0	M 0 0	M 0 0	M 0 0
Cost[a]	H 1 1	L 0 0	H 1 1	M 0 0
Accessibility	M 0 0	H 1 1	M 1 1	L 0 0
Near destination	M 0 0	H 1 1	M 0 0	M 0 0
Travel time	M 0 1	M 0 0	M 0 1	M– 0 0
Frequency[a]	M 0 0	H 1 1	M 1 1	L 0 0
Lack of stress	M 1 1	L 0 0	M 0 0	M 0 0

[a]Criteria considered the most important determinants of competitive position. Short Distance PABT's share of the trans-Hudson market is 58 percent. Bold codes indicate areas of clear current and future superiority.

coverage, they were placed in NJ Transit's matrix represented by a circle whose size indicated its share of total bus or rail ridership. As shown in Figure 5, Short Distance PABT ridership is 6 percent of NJ Transit's total bus ridership, so it is represented by a circle that is half the size of the Bergen, Passaic, Middlesex, Union PABT that carries 13 percent of NJ Transit's bus riders. The size of the circle could also be based on deficit per passenger.

FORMULATION OF STRATEGY

After confirming that the services were correctly placed in the matrix according to the criteria ratings, management reviewed the matrix to determine if NJ Transit's current service mix was satisfactory. Service mix is simply the pattern created by the services depicted on the matrix. In general, NJ Transit's service mix was considered acceptable because it was a strong competitor in both attractive and unattractive markets. In some markets, however, NJ Transit's competitive position was weak, suggesting an inefficient use of public resources. In these cases, management had to decide whether to develop strategies to improve market share, thus changing the service's location in the matrix, or to exit the market gracefully.

NJ Transit management reviewed the market attractiveness and competitive position criteria ratings assigned to the individual route groups and rail lines and formulated strategies for improving or maintaining ratings. To improve the Short Distance PABT's revenue-to-cost ratio and thus overall market attractiveness, for example, management proposed using articulated buses. To improve the Short Distance PABT's reliability, travel time, and nearness to destination, and thus competitive position, management proposed working for a second express bus lane in the Lincoln Tunnel, preferential bus lanes in Manhattan to allow NJ Transit buses to serve the East Side, and other remedies.

CRITIQUE AND CONCLUSIONS

After researching available matrices and strategic approaches and finding them unsatisfactory for the transit industry, NJ Transit developed its own approach based on the MacMillan and AT&T processes. NJ Transit used the overall matrix concept developed by MacMillan. Many of MacMillan's criteria, for assessing both the market attractiveness and the competitive position of a service, were appropriate because they acknowledge the importance of political support and quality of service to a nonprofit agency. Nevertheless, some of the criteria were revised to make the process more appropriate for transit.

NJ Transit's approach evaluates the attractiveness of serving a market and the performance of its services relative to those of other providers. Management first analyzed its services collectively and considered which services to maintain or surrender to achieve the desired portfolio. Then the agency formulated specific strategies to capitalize on its strengths and avoid or eliminate its weaknesses. The process inherently encourages a cost-effective distribution of public resources.

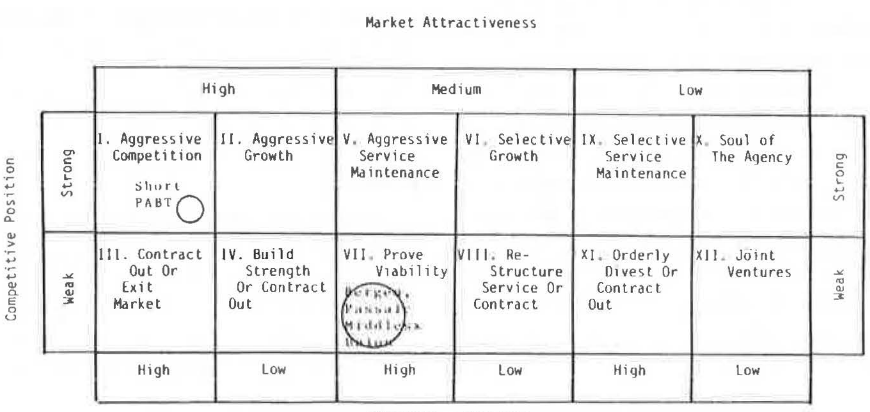


FIGURE 5 Completed NJ Transit matrix.

Other agencies should be aware, however, of several aspects of the process that may hinder its completion. First of all, this process takes a strong commitment of time and energy from the general managers and other top management. In addition, a thorough evaluation of services requires accurate and detailed data. NJ Transit is fortunate to possess some of these data; however, in the future the agency will be supplementing them with more sophisticated market research information. Finally, the process described in this paper is only a part of the strategic planning process. To complete the process, the strategies developed must be reconciled with the available financial resources. The next step is to place financial and political constraints on the process to force choices among the various strategies.

Despite these difficulties, NJ Transit found the process invaluable because it provided a formal structure within which to analyze both the markets and the attractiveness of the available travel options. For the first time, NJ Transit management was able to systematically assess the performance of its own bus and rail operations compared with that of the competition in specific markets. As a result of determining its current and desired service mix, the agency is allocating resources to maintain or improve services in all markets where it is a strong competitor. If it is determined that NJ Transit cannot become a strong competitor in other markets, some services will be contracted out to private operators or eliminated altogether.

REFERENCES

1. G. A. Steiner. *Strategic Planning: What Every Manager Must Know.* The Free Press, New York, 1979.
2. K. E. Stein-Hudson and B. D. McDowell. Applying Strategic Planning in the Transportation Sector. *TR News*, July–Aug. 1985, pp. 20–24.
3. *Developing Competitive Strategies Workshop.* The Organization Effectiveness Group, AT&T, 1984.
4. S. K. Johnson. Evaluating the Corporate Portfolio. In *Handbook of Business Strategy* (W. D. Guth, ed.), Warren and Lamont, Boston, Mass., 1985, Chapter 6, pp. 1–17.
5. I. C. MacMillan. Competitive Strategies for Not-For-Profit Agencies. In *Advances in Strategic Management* (R. Lamb, ed.), JAI Press, Stamford, Conn., 1983.

Publication of this paper sponsored by Committee on Strategic Management.

Strategic Management in a Crisis-Oriented Environment

MICHAEL D. MEYER

Strategic management is a concept that has been applied for many years in the private sector. Only recently have public-sector transportation agencies become interested in strategic management. In this paper the literature on strategic planning and strategic management is reviewed and a definition of strategic management is offered. A strategic management application in one state highway agency is examined. The results of this study are used to make observations about characteristics of successful strategic management. It is concluded that a strategic management process is an important managerial planning tool for dealing with a rapidly changing policy environment such as that facing transportation agencies.

A basic tenet of effective organization management is that managers should play an important role in determining the strategic direction of their agency. Nowhere is this function more important than in state transportation agencies. For many years such agencies focused their resources on the design and construction of new facilities. Today, however, the environment of transportation is changing dramatically. Many transportation agencies are now more concerned with maintaining the existing system than with building new facilities. In addition, changes are being proposed in the federally aided highway and transit programs that could significantly affect the way transportation agencies do business. In addition, many transportation agencies will lose much of their professional staff to retirement during the next several years. These and other factors indicate a need for a systematic process for assessing the strategies available to an agency for dealing with future threats and opportunities, and for implementing the most effective strategies. Such a process is called strategic management. The purpose of this paper is to identify the key characteristics of a strategic management process and to examine the critical dimensions of implementing such a process in a public-sector organization. To accomplish this, the literature on strategic planning and strategic management is examined. Because strategic planning forms the basis for strategic management, some time is spent in the first section discussing the key characteristics of strategic planning. Then one strategic management application in a state highway agency is examined, and the results of this case study are used to draw conclusions on the substance and manner of implementing strategic management in public-sector transportation agencies.

Bureau of Transportation Planning and Development, Massachusetts Department of Public Works, 10 Park Plaza, Room 4150, Transportation Building, Boston, Mass. 02116.

PERSPECTIVES ON STRATEGIC PLANNING AND MANAGEMENT

Strategic planning and strategic management have been applied in the private sector for many years. Perhaps the best discussion of these concepts, and of how they relate to one another, is given by Ansoff (*1*). As shown in Figure 1, Ansoff places the first major adoption of corporate strategic planning in the United States during the late 1950s. Such planning was initiated to deal with uncertainty relating to foreign competition, decline of some major industries, rapid technological advances, and product diversification. It was not until the late 1970s, however, that the concept of strategic management first appeared. It was at this point that corporate managers realized that the planning of strategy and, more important, strategy implementation, could not be divorced from the planning and management of an organization's capability.

Throughout this period, the business and management literature was filled with technical articles on how to conduct strategic planning and how to manage strategically. Some important contributions were made by Anthony (2) who developed a framework for examining planning and control systems within an organization. Anthony defined strategic planning as the process of deciding on organizational objectives; on changes in these objectives; on the resources used to attain these objectives; and on the policies that are to govern the acquisition, use, and disposition of these resources.

Ackoff, in analyzing the concept of corporate planning, identified five major parts of a corporate plan and, hence, of the phases of a corporate planning process (*3*). These phases involve specification of objectives and goals, selection of policies and programs to achieve these objectives, determination of the resources needed to implement these actions, design of a decision-making framework to carry out the plan, and establishment of a monitoring mechanism to detect and prevent errors in plan implementation. The value of corporate planning to managers was also considered by Ackoff to lie more in their participation in the process than in their use of the resulting document.

As the concepts of corporate and strategic planning became of greater interest to many top managers, the literature on these topics expanded rapidly, with many authors providing different definitions of the concepts. Drucker (*4*), for example, viewed strategic planning as an entrepreneurial skill and as a continuous process of making present decisions with an awareness of future opportunities and consequences. Andrews (*5*) similarly viewed strategic planning as establishing the pattern of major

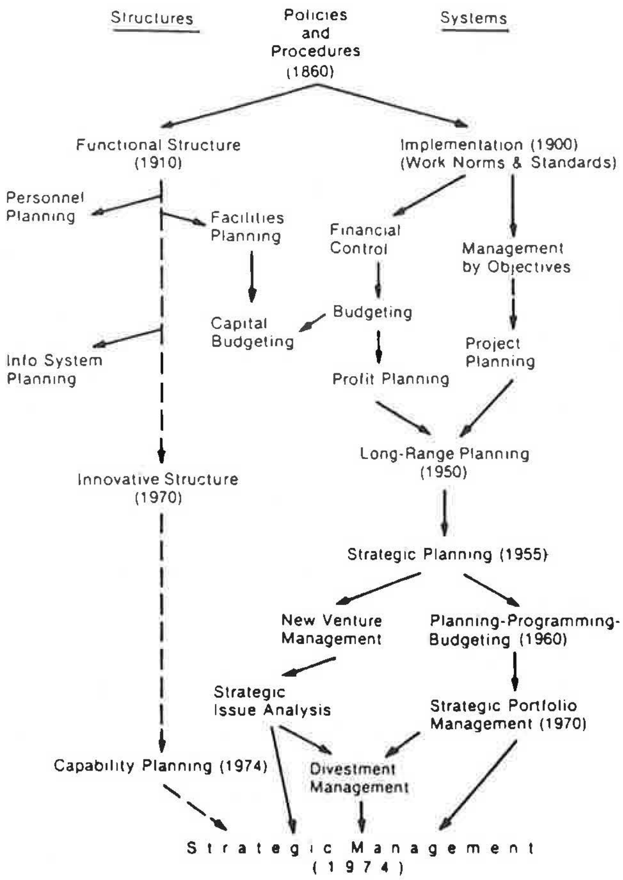


FIGURE 1 History of strategic planning and strategic management.

objectives, purposes, or goals and identifying the policies and plans for achieving these goals. His framework included an assessment of opportunities in the firm's environment; the integration of these opportunities with the technical, financial, managerial, and personnel resources of the organization; and a consideration of this integration with agency mandates and goals.

Perhaps the most thoughtful examination of strategic planning was conducted by Steiner (*6*). Acknowledging that the field of inquiry had been flooded with different definitions of strategic planning, Steiner noted that strategic planning should be approached from four points of view.

First, planning by definition deals with the futurity of current decisions. Strategic planning is thus "the systematic identification of opportunities and threats that lie in the future, which in combination with other relevant data provide a basis for a company's making better current decisions to exploit the opportunities and to avoid the threats."

Second, strategic planning is a systematic, continuous process of setting and validating organizational goals, defining strategies and programs to achieve these goals, and developing detailed plans to implement strategic decisions. Three of the major outputs of a strategic planning process are (a) a statement of organizational goals and objectives, (b) a plan that outlines the evolution of the agency over specific time periods, and (c) work programs (usually at division levels) that establish the direction for organizational work units and act as a means of monitoring progress toward desired agency performance.

Third, strategic planning is as much an organizational philosophy as it is a planning process. Top management must place importance on the activities associated with strategic planning and on the results of the process. Members of the organization must view strategic planning as an effort worth undertaking.

Finally, strategic planning provides an important link among operations plans, medium-range programs, and budgets. This link is critical in coordinating the many organizational activities that can play an influential role in helping top management achieve strategic objectives. An effective coordination effort also provide credibility to the strategic planning process.

Even as these and other authors argued the merits of strategic planning, others began questioning the effectiveness of the strategic planning process. Ansoff et al. (*7*) argued that strategic managerial issues were too complex to be handled by strategic planning. Others challenged the basic assumptions of strategic planning and argued that a more broad-based approach was needed to address strategic organizational issues (*8–10*). A major criticism of strategic planning was that it did not often have a close relationship with the actual implementation of the options defined in the strategic planning process; that is, making decisions about changes in organizational structure, allocation of personnel and budget, and monitoring the effectiveness of strategy implementation. The link between strategic planning and strategy implementation resulted in the process called strategic management.

The foregoing discussion provides a brief description of how strategic planning and strategic management have evolved in the private sector. Application of these concepts in the public sector, and specifically in transportation, has been a matter of adapting, where possible, approaches and techniques from the private sector. This application, however, has not occurred without debate about its usefulness in the public-sector environment. Steiner argued that public-sector application of strategic planning was limited to certain types of policy issues because of the political nature of that environment (*6*). Rondinelli (*11*) similarly argued that public-sector planning involves political conflict and resolution, often with no clearly defined criteria for evaluation of alternatives. Thus, he concluded that the more structured corporate-sector planning process cannot be applied directly to the public sector. Others, however, have concluded that there are sufficient similarities between the two sectors to make it possible to use strategic planning with some success in the public sector (*12–16*). A recent review of public-sector strategic planning concluded that, although its use is relatively new, experience to date has indicated that public-sector managers can derive benefits from such a process (*17*).

In the transportation sector, strategic planning and management have only recently received serious attention. A 1983 review of strategic planning in transportation agencies found that some form of strategic planning existed in several Canadian agencies and in a few state transportation and port authority organizations in the United States (*18*). Given the rapidly changing environment of transportation agencies, this review concluded that strategic planning would be a valuable tool for transportation managers. Indeed, much of the interest in strategic planning in the early 1980s was found in public transit agencies that were facing major uncertainty because of potential future cutbacks in federal funding. Several authors noted the importance of a strategic planning process for addressing this uncertainty (*19–21*). Overall, however, the literature on strategic planning and management in transportation has been sparse.

In summary, private-sector applications of strategic planning and strategic management concepts have far outnumbered those found in the public sector. Even so, there has been a growing interest in the transportation sector in applying these concepts to better anticipate the problems and opportunities of a rapidly changing world. Although some authors have argued that there are substantial differences between public- and private-sector applications, there is growing evidence that a strategic management process is a useful tool for transportation managers to employ in dealing with a rapidly changing policy environment.

For the purposes of this paper, strategic management will be defined simply as the process by which managers understand organizational goals, examine the future threats to and opportunities for an organization, identify strategies for dealing with these threats and opportunities, change organizational capability to implement these strategies, and continually monitor the entire process to provide managerial direction and support for accomplishing the strategic management objectives. This definition will be used in the following case study to illustrate the implementation of strategic management in one state highway agency.

STRATEGIC MANAGEMENT IN A STATE HIGHWAY AGENCY: A CASE STUDY

The Massachusetts Department of Public Works (DPW) is the agency responsible for the planning, design, construction, and maintenance of the state highway system in Massachusetts. For many years, the DPW used its strong professional capability to design and build this system. However, by the late 1960s and early 1970s, the department was embroiled in numerous conflicts over the future direction of the state highway program. In the Boston area, for example, a multiyear study of transportation issues resulted in a gubernatorial moratorium on most major highway construction.

By the late 1970s, it had become clear to top DPW management that significant changes were likely to face the DPW in the coming years. A meeting of 37 top managers held in 1978 to discuss the problems facing the DPW identified some of the key issues facing the agency at that time:

- There will be a smaller number of "big-build" Interstate-type projects in the future,
- Maintenance of the highway system will become of increasing concern,
- The state aid program to local communities will become less important because of insufficient resources to administer the program,
- Coordination with other agencies will become more important because of environmental and intermodal coordination issues, and
- Federal and state initiatives on affirmative action and minority enterprises will require increased attention on the part of DPW managers.

Before top management could address these problems, a new governor was elected and several top managers were replaced. Four years later, however, a review of the DPW by the FHWA indicated that the problems had become worse during the intervening 4 years. The FHWA predicted that, in its then current condition, the DPW would not survive for more than 2 years. This prediction was based on several factors. In 1980, the DPW had nearly 4,000 employees. By 1982 this number had been reduced to 2,900 with approximately 1,000 of these receiving demotions. This staff reduction, caused by budget cutbacks resulting from a referendum on tax limitations, created a serious morale problem in the DPW. The average age of DPW employees had also increased to 57 years because those fired were most often the youngest. In addition, the number of construction projects advertised had reached a modern era low in 1982.

When a new governor was elected in 1982, he faced a serious problem in rebuilding the DPW. The new top management looked at ways to structure this rebuilding effort and began a strategic management effort that had several successes and some failures. These successes and failures will be examined in the rest of this case study.

For purposes of presentation, the analysis of the DPW's strategic management effort will be divided into the strategic management components defined in the last section.

Understanding Organizational Goals

The first task in the strategic management effort was to better understand the organizational goals of the DPW. The top 20 managers were asked to define these goals on paper. Much to the surprise of this group, there were clear differences of opinion on what these goals were. The engineers who had been in the DPW for many years focused the goals on the engineering, construction, and maintenance of the state highway system. The new managers, some nonengineers, defined the goals more broadly and related them to the DPW's role in the entire transportation system. After numerous discussions, the head of the agency drafted a mission statement that served as the basis for the mission statement that was eventually adopted:

- To allow and promote the mobility of people and goods in Massachusetts through the sound development, efficient operation, and reliable maintenance of a safe and attractive highway system and through coordination of that system with other transportation agencies to form a coherent public transportation network for all users;
- To administer highway capital programs so that transportation goals are met to promote economic welfare, with maximum benefit to the physical and social environments; and
- To assist local governments in improving their highway networks in both urban and rural areas.

Examining Future Threats and Opportunities

The next task in strategic management is to examine future threats and opportunities, in actuality the strategic planning component of strategic management. Although this task is critical for establishing the strategic agenda of top management, there is little consensus in the literature on how such a task should be accomplished. The "environmental analysis" or "environmental scanning" procedure used in private-sector applications has employed techniques ranging from the Delphi method to economic input-output models (22). Such a level of sophistication, however, was considered unnecessary for the DPW effort.

The top 20 managers were asked to identify the critical problems that would face the DPW in the next 5 years. More than 30 different problem statements were received and categorized in six major areas: project selection and development, organizational structure, personnel, funding, maintenance and bridge rehabilitation, and public image. A matrix was then formed with these six issues as rows and four questions as columns (Figure 2). The four questions were

- Are there future threats that will exacerbate this problem?
- Are there future opportunities that could expedite solution of the problem?
- What actions are currently being undertaken?
- What actions should be undertaken?

The last column in the matrix, once filled in, would thus serve as the strategic planning agenda for top management.

Over a period of 2 months the strategic management committee met several times to complete the matrix shown in Figure 2. Examples of the way two issues were defined follow.

1. Issue 1: Project selection and development
   Future threats
   - There are too many projects in the pipeline given limited funds.
   - The Southeast Expressway and Central Artery projects will consume much of the department's resources. The department might not be able to support other projects.
   - The passage of new legislation will likely raise expectations of department project delivery and might even require that commitments be made to assure passage.
   - Political pressures will be brought on the department to deliver.
   - Continuing economic development pressures will require some action by the department to provided needed infrastructure.
   - If the legislature creates an independent infrastructure finance bank, this could promote uncoordinated highway project development and could even cause the department to lose some discretion in highway decisions.
   - The organization of the department, if it remains the same, will not provide for effective project development.

   Future opportunities
   - New legislation will provide added money and personnel to rebuild the department project implementation capability.
   - The trade-in of Interstate projects will provide additional funds for projects.
   - The passage of time will see a steady deterioration of the highway system. This deterioration could create a useful justification for added funds.
   - Political influence could be orchestrated to promote additional funding.

   Current actions
   - The Interstate substitution transfer.
   - Bond legislation.
   - The development of an infrastructure slide show.

   Future actions
   - Determine viable criteria for the selection of projects and stick to them.
   - Develop a strategy to educate the public and legislature about the project development process; that is, clearly define the process.
   - Develop a realistic 5-year program and a process to examine 10-year needs.
   - Upgrade project information system to allow on-line information on all of the department's projects.
   - Focus public attention on the bridge problem and on the department's response.
   - Develop a strategy for increasing funds in capital and operations budget, perhaps exploring alternative sources of dedicated funds.
   - Conduct a study of the project selection and development process with explicit consideration given to the "batching" of projects at key decision points.
2. Issue 6: Perception

	Future Threats	Future Opportunities	Current Actions	Future Actions
Project Selection and Development				
Organization Structure				
Personnel				
Funding				
Maintenance and Bridge Rehab.				
Public Image				

FIGURE 2 Strategic planning matrix.

Future threats
- Possible increase in infrastructure failures (e.g., bridge collapses).
- The Southeast Expressway reconstruction will create a terrible image of the DPW.
- An overcommitment of projects will clearly strain the DPW's credibility if it cannot deliver.
- Cutbacks in the state-aid program and the DPW's role could alarm cities and towns.

Future opportunities
- The move to the new building could be used to paint a new image of the DPW.
- Good publicity on the Southeast Expressway project could help the DPW image (although this will likely be a no-win situation).

Current actions
- Interface between local citizens and engineers.
- Major spring cleanup.
- Tourist information program.

Future actions
- Conduct a poll to determine DPW's image.
- Develop a media strategy to associate DPW with "public interest" topics.
- Conduct workshops in cities and towns.
- Devote considerable effort to planning mitigating actions for the reconstruction of the Southeast Expressway.

This strategic planning effort provided a systematic process for identifying key strategies that should be undertaken to prepare the DPW for the future. Several of these strategies were implemented. For example, considerable effort was spent on planning the mitigation plan for the reconstruction of a major Boston expressway. This effort, identified as both a threat and an opportunity in the matrix, resulted in substantial favorable publicity for the department and is widely considered to have enhanced its image. Other strategies were not implemented, however, mainly for one significant reason: the day-to-day demands on top managers required almost all of their attention. When decisions needed to be made on today's problems, there was little time to consider actions that would have an impact several years hence. And there was no direction from upper management that managers should, in effect, make the time.

Many of the strategic initiatives identified in this task were thus not implemented because of competing demands for managerial attention. This was not true for one of the most important issues, however.

Changing Organizational Capability

Given limited resources and time, the head of the agency decided that an assessment of the DPW's organizational structure and of its capability to handle future work should receive priority. A consultant was hired to conduct an organizational analysis of the DPW and to recommend changes that were considered necessary. Because this analysis is really a key component of the DPW's strategic management effort, and because its results are being implemented (which will have significant impact on the DPW), some time will be spent here discussing the key characteristics of the organizational assessment, which are shown in Figure 3.

Several of these tasks merit special attention. Because organizational structure and responsibility should reflect the mission and goals of an agency, it is extremely important to begin the assessment with a common understanding of mission, goals, and mandates. Thus, the first task shown in Figure 3 is critical for a successful assessment effort. The organizational analysis task serves as the basis for the entire assessment process. This task is often quite difficult because of the need to define the formal and informal lines of authority and communication. The formal organizational chart does not often reflect what actually occurs in an organization.

The implementation tasks are also a key ingredient to successful organizational change. Not only are recommended changes to the organizational structure important; the identification of equipment needs, human resources, and required changes to legislative and regulatory mandates are as well. Without these implementation tasks, managers might have a difficult time developing a set of specific steps needed to implement organizational changes. The assessment process shown in Figure 3 resulted in the identification of several organizational and staffing issues that needed to be addressed to prepare the DPW for the future. Example issues include

- Reducing the span of control of top managers,
- Strengthening the strategic planning process,
- Strengthening contract management and design functions,
- Revising out-of-date standard operating procedures,
- Increasing use of computer and word processing technology, and
- Dealing with retirement of current staff (44 percent) over the next 5 years.

These issues were then used to develop specific organizational and staffing recommendations that, for the most part, have been adopted (*23*).

Monitoring Strategy Implementation

The DPW has just begun to implement the recommended organizational changes. A top management committee has been formed to develop an overall strategy for this implementation and to monitor the effectiveness of these changes. It is thus too soon to judge the success or failure of this monitoring effort.

Although DPW experience with strategic management is still in its early stages, there has been sufficient exposure to the process to allow several conclusions to be drawn about what is needed for successful strategic management.

STRATEGIC MANAGEMENT IN A TRANSPORTATION AGENCY: LESSONS LEARNED

Although the strategic management process described is related to one specific case, several observations can be made that relate to public-sector strategic management in general. These observations are offered as a reference for further research on the characteristics of successful strategic management in the public sector.

- Review Mission, Goals, Legislative Mandates and Related Information
- Interview Key Management Personnel
- Oversee "Peer Review" of Similar Departments
- Conduct Organizational Analysis
  - Review Current Organization Structure and Staffing Plan
  - Determine Responsibilities and Functions Performed by Each Organizational Element
  - Determine Formal and Informal Lines of Communication
  - Utilize Charting Techniques to Identify Problem Areas, such as:
    - Fragmentation of Functional Responsibility
    - Excessive Span of Control
    - Staffing Imbalances
- Conduct Operations Analysis
  - Review Current Operational Procedures Relating to Maintenance and Construction Programs
  - Identify Opportunities to Improve Program Effectiveness and Efficiency through Improved Operational Practices
    - Maintenance Management
    - Inventory Management
    - Priority-Setting Methods
- Perform Staffing Analysis
  - Review Recent Staffing Trends and Characteristics
  - Review Current Staffing Plan
  - Assess Current Staff Resources Relative to Current Program Needs
  - Assess Current Staff Resources Relative to Expected Program Needs
- Develop Recommended Organizational Structure and Staffing Levels
- Define Phased Organizational Changes to Implement Recommended Organizational Structures
  - Grouping of Department Functions by Organizational Units
  - Revised Lines of Communication
  - Assignment of Program Responsibilities
  - Identify Level of Staffing by Function and Organizational Unit
  - Determine Staffing Needs during Period of Transition
- Oversee Equipment Needs Definition
  - Determine Information Needs
  - Develop Management Plan Elements
    - Maintenance Management
    - Equipment Management
    - Performance Monitoring and Evaluation
- Prepare Human Resources Action Plan
  - Career Development
  - Training Programs
  - Recruitment Policies
- Confirm Funding Assumptions and Requirements
  - Confirm Funding Requirements and Sources
  - Define Financial Management Issues
- Identify Applicable Legislative and Regulatory Changes
  - Identify Necessary Amendments to Existing State Legislation
  - Identify New Legislative Authority
  - Identify New or Amended Administrative Regulations
- Develop Master Implementation Schedule
  - Define Timetable for Transition to Recommended Organizational Structure
  - Identify Organizational Changes, Staffing Levels and Program Requirements Associated with Each Schedule Milestone

FIGURE 3 Key tasks of organizational assessment.

Impact of the Organizational Environment

The DPW is primarily an implementing agency, with a strong organizational culture oriented toward short-term action. Because of this orientation, agency action (and thus managerial attention) is often heavily influenced by events that occur outside the agency. For example, a truck accident on a major highway can demand the attention of several managers for an entire day. Other examples of such events include public controversy over important projects, media attention to the agency or agency projects, communication from influential political leaders, and legislative initiatives that need to be guided through the political process.

Because transportation agencies have such an important impact on the efficient operation of the economy of a state or city, it is not surprising that the organizational environment puts pressure on agency managers to focus their attention on the short term. However, this creates a serious challenge to successful implementation of a strategic management process.

Upper Management Commitment

Almost every book and article written on strategic planning and management observes that successful efforts require top management commitment to the process. The DPW case once again illustrates this observation. Because of the demands placed on managerial time, upper management must make it clear that assessing the strategic direction of an agency is an important task for managers, and that time must be found to accomplish this task. Without such direction, managers will focus their attention on those issues that confront them day to day.

Process Flexibility

One of the major criticisms of strategic management in the private sector has been its rigid structure, often dictated by the demands of the analytical procedures used for strategic planning. Such a rigid structure can stifle the creativity that is necessary to undertake strategic management. In the DPW case, the managers themselves defined what the structure was to be and did not allow the process to overly influence the results. It is important to note that upper management did not delegate the strategic planning function to the staff. Strategic planning was done by the managers themselves.

Strategy Implementation

Successful strategic management provides a strong link between planning and implementation. In the DPW case, this link was seen in the organizational assessment that resulted in recommendations for specific organizational changes. A continuous strategic management effort would not necessarily result in periodic changes to the organizational structure, unless they were warranted. Such an effort, however, should provide a strong link between the planning component and the resource allocation functions (i.e., budget and personnel) in an organization. If this link does not exist, the strategic management effort may not accomplish its objectives.

Human Resources

Assessing an organization's capability includes an examination not only of organizational structure but also of the skills and characteristics of the staff. This has been one of the major issues left out of most strategic management efforts. For transportation agencies, many of which are facing large turnovers in professional staff, the human resources issue could become a critical component of strategic management.

Outside Help

Although agency managers must be the most active participants in strategic management, it is often worthwhile to bring in expertise from outside the organization to help with the process. In the DPW case, outside help provided an organizational analysis capability that was not available in the agency. It is important to note that the consultants who conducted the organizational assessment did so in strong coordination with top management officials of the DPW. The consultants acted as a catalyst in helping DPW managers think about how the agency should be organized.

CONCLUSIONS

The Massachusetts DPW case illustrates how strategic management has been applied in one situation. Clearly, such an effort would be structured differently in other contexts, with varying degrees of manager participation and levels of analytical sophistication. However, the DPW case does suggest that, especially for an agency that has an implementation orientation, some form of strategic management is necessary to focus managerial attention on the organization's future.

Strategic management consists of four major steps: understanding the organization's goals, identifying key changes likely to occur in the organization's environment, assessing an organization's capability for dealing with these changes, and establishing an institutional mechanism for monitoring the strategic management process. In this definition, strategic management is not only a planning tool but also an important management function. Given the rapidly changing policy environment facing transportation agencies, a strategic management process is critical for focusing the attention of managers on the likely implications of these changes.

Strategic management would appear to be most effective when upper management is committed to the process and has so informed agency managers, when the process is sufficiently flexible to allow wide-ranging participation of agency managers, when the strategic planning component is clearly related to implementation strategies such as budget and personnel allocation, and when organizational capability is viewed from a human resources perspective as well as from a structural point of view.

REFERENCES

1. H. I. Ansoff. The State of Practice in Planning Systems. *Sloan Management Review*, Winter 1977.
2. R. N. Anthony. *Planning and Control Systems: A Framework for Analysis*. Graduate School of Business Administration, Harvard University, Cambridge, Mass., 1965.

3. R. L. Ackoff. *A Concept of Corporate Planning*. Wiley-Interscience, New York, 1970.
4. P. F. Drucker. *Management*. Harper & Row, New York, 1974.
5. S. K. Andrews. *The Concept of Corporate Strategy*. R. D. Irwin, Homewood, Ill., 1980.
6. G. A. Steiner. *Strategic Planning: What Every Manager Must Know*. The Free Press, New York, 1979.
7. H. I. Ansoff, R. P. Declerck, and R. L. Hayes, eds. *From Strategic Planning to Strategic Management*. Wiley-Interscience, New York, 1976.
8. R. O. Mason and I. I. Mitroff. *Challenging Strategic Planning Assumptions*. Wiley-Interscience, New York, 1981.
9. S. C. Sufrin and G. S. Odiorne. The New Strategic Planning Boom: Hope for the Future or a Bureaucratic Exercise? *Managerial Planning*, Jan.–Feb. 1984.
10. C. Stubbart. Why We Need a Revolution in Strategic Planning. *Long Range Planning*, Vol. 17, 1984.
11. D. A. Rondinelli. Public Planning and Political Strategy. *Long Range Planning*, Vol. 16, 1983.
12. J. M. Bryson and W. D. Roering. A Review and Critique of the Major Approaches to Public Sector Strategic Planning. Presented at the American Planning Association National Conference, Los Angeles, Calif., 1986.
13. M. A. Murray. Comparing Public and Private Management: An Exploratory Essay. *Public Administration Review*, Vol. 35, 1975.
14. W. R. Dodge and D. C. Eadie. Strategic Planning: An Approach to Launching New Initiatives in an Era of Retrenchment. *Management Information Service*, Vol. 14, 1982.
15. J. B. Olsen and D. C. Eadie. *The Game Plan: Governance With Foresight*. The Council of State Planning Agencies, Washington, D.C., 1982.
16. R. J. Vaughan and R. Pollard. *Rebuilding America, Planning and Managing Public Works in the 1980's*. The Council of State Planning Agencies, Washington, D.C., 1984.
17. C. Bloom. Strategic Planning in the Public Sector. *Journal of Planning Literature*, Vol. 1, No. 2, Spring 1986.
18. M. D. Meyer. Strategic Planning in Response to Environmental Change. *Transportation Quarterly*, April 1983.
19. D. R. Schwartz. Strategic Planning for Transit Properties. Presented at the American Public Transit Association Meeting, Denver, Colo., 1982.
20. M. G. Ferreri. Strategic Planning for Transit. Presented at the American Society of Civil Engineers Annual Convention, New Orleans, La., Oct. 1982.
21. M. D. Meyer. Strategic Planning in Response to Environmental Change: The Transit Case. *Selected Papers of the Planning and Policy Committee*, American Public Transit Association, 1983.
22. P. De Vasconcellos Filho. Environmental Analysis for Strategic Planning. *Managerial Planning*, Jan.–Feb. 1984.
23. Peat, Marwick, Mitchell & Co. *Proposed Organizational Structure, Functional Responsibilities and Staffing Levels*. Massachusetts Department of Public Works, Boston, June 1986.

Publication of this paper sponsored by Committee on Strategic Management.

Heuristic Decision Framework for Upgrading Highway Weight Limits

YORGOS J. STEPHANEDES, J. ZIOTAS, AND S. ARORA

A heuristic decision framework is developed for obtaining a regional road development program that optimizes the net benefits of the projects in the program and meets a specified budget constraint. Because a regional network serves a considerable number of plants and markets and consists of a large number of links, the benefits that result from improving a single link are almost never immediately realized. In the case of a program to upgrade highway weight limits, a benefit is realized only when the minimum load limit along a travel route is raised. The heuristic algorithm addresses this special constraint and determines optimal road development plans for various budget levels. Although this analysis concentrates on selecting projects that upgrade the weight limits on state highways, the methodology is also applicable to other types of highway project selection.

The method described in this paper was developed as part of a larger research project that seeks to identify the possible interactions between state transportation expenditures and economic development. The issue of determining the existence and size of these interactions is addressed elsewhere (*1*, *2*). This paper deals with the ways expected project benefits and costs (including any economic impacts) can be considered in highway project selection. In particular, a framework is developed for obtaining a road development program that optimizes the net benefits of the projects in that program while meeting a budget constraint. Although the analysis is focused on selecting projects that deal with changing the weight limits on state highways, the method is formulated in a general manner so that it could be applied to other types of highway project selection.

Because a regional highway network serves a considerable number of plants and markets and consists of a large number of links, the benefits that result from improving a single link of the network are almost never immediately realized. More specifically, the net benefit impacts are not fully realized until the reactions of shippers and carriers to route improvements have taken place and any economies or cost savings resulting from such changes are worked into pricing structures and production levels such that consumer-producer relationships are affected. Alternatively, network links could be upgraded in sets so that the lowest construction costs resulted in the maximum realizable benefits. In the case of weight limits, a benefit is realized only when the minimum load limit along a route is raised (where the minimum load limit along a route is equal to the maximum allowable load on that route), and it is this special feature (constraint) that makes the problem interesting. Because the problem does not appear to be amenable to an obvious dynamic programming formulation, a heuristic algorithm that determines optimal development plans for various budget levels was developed. The heuristic algorithm is based on complete enumeration, a technique that is appropriate for reasonably sized problems such as the one studied here. The analysis is applied to transportation benefit and cost data on the forest industries and the highway system in northeastern Minnesota. In this application, changes in weight restrictions and upgrading and expanding year-round 10-ton state routes are expected to affect transport cost per mile and direct yearly benefits. For instance, assuming constant demand and supply between origins and destinations, the direct benefits depend on the number of trips saved, the transportation unit cost, the length of trips, and the annual time period in which the benefits occur. If the shipping patterns of forest industry products remain consistent with previous shipping patterns and the number of shipments is not reduced by the closing of forest plants, the impact of upgrading a forest product route may be significant.

Given a set of benefit and cost criteria, it may be possible to estimate the impacts resulting from upgrading a highway network. However, the challenge is to employ the results of such an impact analysis to establish and execute a systematic process that will lead to the optimal distribution of available funds to the network. Before the methodology that was developed to aid project selection and assure an optimal fund distribution in a road network is outlined, a background and review of the subject and other major project selection studies are presented.

BACKGROUND

"The demand for highway improvements is increasing much more rapidly than funds are becoming available. Consequently, all jurisdictions in a state feel cheated. . . . Perhaps the best that could be hoped for is that everyone would feel equally cheated" (*3*). Decisions on when, where, and what type of improvements to make are some of the most important tasks faced by transportation agencies at all levels of government. But before decisions can be made, adequate criteria and standards representing the efficiency, effectiveness, and equity aspects of a project need to be established. Techniques are, then, required to assist in the evaluation of options for decision

Y. J. Stephanedes and J. Ziotas, Department of Civil and Mineral Engineering, S. Arora, Department of Industrial Engineering and Operations Research, University of Minnesota, 500 Pillsbury Drive, S.E., Minneapolis, Minn. 55455.

making. Also needed are methodologies to set priorities in programming of projects in a limited financial environment (*4*).

Substantial work has been done on the criteria employed and the nature of the highway programming process in the various states (*3–5*). Highway cost allocation methods (*6–8*) and maintenance programs (*9*) have also been described in detail. Computer-based methods (*4*) and technical procedures have been introduced, but also criticized. For instance, such procedures often take so long to apply that funding decisions must be made without the benefit of those procedures (*3*); yet existing benefit-cost investment rules have been found naive and the need for more sophisticated rules has been identified (*10*). A major criticism of the recent U.S. highway cost allocation approach (*8*) is that it is based on expenditures, not costs. Any expenditures that are incurred in a particular year are allocated to the traffic of that year even though the benefits arising from the investments are realized over a longer period. Such an approach neglects the indivisibilities that are necessarily involved in the provision of highway infrastructure and the resultant excess of capacity and cost (*6*). A similar problem arises in upgrading a road network by raising weight limits; in such a case, a benefit can be realized only when a whole travel route is upgraded. In this paper, this nonlinear problem is considered and alternative methods for addressing it are presented.

A small number of investment programming studies have developed highway programming methods based on estimated costs and benefits to highway users (e.g., operating costs, travel time) and nonusers (e.g., governmental costs). Typically these are combinatorial optimization methods such as linear and dynamic programming and branch-and-bound techniques. Bergendahl (*11*), for instance, employed a combination of linear and dynamic programming to determine the optimal size and time for investments in new highway links in southern Sweden. He decomposed the problem into a set of network problems in which each network represented the road system in a different phase of development. The road network was assumed fixed in 5-year periods and investments could be undertaken only at these time intervals. The optimal investment between periods was determined by minimizing current operating cost, where link cost was a convex monotone increasing function of traffic flow.

The Dutch Integral Transportation Study (*12*) devised a method for minimizing the investments and user costs in the Dutch network from 1970 to 2000. To minimize computation time, the method decomposed the original problem into smaller networks, optimized through linear programming. These results were used in the master problem with a stepwise capacity-restraint assignment according to a least marginal objective and a descriptive route choice model and led to the minimization of a social cost objective function.

A third example of an investment programming method is the Highway Investment Analysis Package (*13*), which uses microeconomic theory to analyze individual roadway sections and limited networks of sections specified by their physical, traffic, and operational characteristics. It is composed of four computer modules that do not guarantee a globally optimal solution but produce efficient solutions that satisfy all constraints.

Further, Schnuerer (*14*) studied the optimization of road investments in the province of Salzburg, Austria, based on travel times and using a dynamic programming model. He examined costs of road improvements that arise from different terrain conditions and travel times that result from alternative speed-design standards that vary within each link of a route. Although the links of a route differ in their construction cost and design-speed functions, he assumes a convex monotone increasing function between costs and design speed to be valid for all of the links. Frequently, however, a link belongs to several routes and, therefore, the reconstruction requirements for that link are determined by several standards (e.g., routes with 50 and 60 mph speed limits), a problem the author does not indicate how he addressed. Combinatorial optimization methods, such as the ones reviewed, could in principle also be used to address the problem under study (i.e., optimization of road investments). In particular, investments for upgrading and expanding year-round 10-ton state routes should be optimized on economic criteria determined by the needs of an economy (e.g., the Arrowhead region of northeastern Minnesota). But these economic criteria could also reflect social factors. For example, a highway improvement could take place even if the dollar benefit is small as long as the revitalization of a disadvantaged section of the region is significant in terms of employment or stabilization of declining towns.

The realized economic benefits of road investments are quantified through transportation cost reductions. However, the benefits vary among the industries of an economic sector because the method of transportation cost payment varies from one industry to another. In the forest industry, for instance, an examination of alternative payment structures is necessary because changes in factors affecting the transportation cost determine different schemes of benefits for the shippers and the freight-carrying companies. Some shippers pay the freight-carriers a flat rate for the movement of their products. Others, contracting with independent truckers, pay (a) by the loaded miles, (b) by the running mile, or (c) by the loaded miles with an additional hourly rate for time spent at the truck terminal. Shippers who lease trucks pay according to a lease agreement. In the first payment alternative, transportation cost reductions are a benefit to the carriers; in the rest of the cases the benefits are enjoyed to a larger extent by the shippers. In the next section, a heuristic procedure is developed to solve the problem of combining maximum realizable economic benefits, which result from the alleviation of weight restrictions or other road improvements, with minimal incurred construction costs expended for the upgrading of the network links. To be sure, both benefits and costs are amortized over the time horizon appropriate for each project. The principles of the heuristic optimization procedure are illustrated with an example.

PROBLEM AND METHOD

In this section, a method for obtaining a road development program that optimizes project net benefits under a budget constraint is developed. As can every combinatorial optimization problem, the current problem can in principle be solved by "exact" techniques (e.g., tree search, branch and bound), but these techniques frequently require computation times that grow faster than polynomially with the size of the problem and get out of control with excessively large problems. Although most combinatorial optimization problems can also be transformed into integer programming models, the disadvantage of that approach is that the mathematical techniques for treating such models are generally inefficient (*15*) although certain efficient heuristic techniques (e.g., Lagrangian relaxation heuristics) have been suggested [for a detailed review of the literature see Crowder et al. (*16*) and Magnanti and Wong (*17*)].

When exact mathematical techniques or integer programming models are inefficient in solving a problem, there are two ways to overcome the dilemma. Either the problem has to be modified, by relaxing the elements causing the algorithmic difficulties, or heuristic procedures must replace the exact mathematical techniques. It is usually advantageous to leave the problem unchanged and develop heuristic procedures (i.e., "systematic" procedures that are precisely defined and, therefore, can be programmed for a computer).

In the problem under study, exact techniques (e.g., linear or dynamic programming) are not applicable because the principle of optimality does not hold. This is evident because, first, minimization of total construction costs and maximization of benefit-to-cost ratios or net benefits may dictate the upgrading of different routes depending on the proposed road construction sequence and, second, different budget constraints should be considered in predicting what is optimum in the process of road investments.

Apart from the linear and dynamic programming methods, no practicable conventional procedure minimizes road improvement costs in a complete road network while considering all possible route combinations simultaneously. Such a simultaneous optimization could not be tackled because of the large number of decision variables and constraints. Thus, in this analysis, a technique is developed for determining the best solution in a stepwise procedure. Although the realizable benefits of each upgraded link depend on the load category of other links, costs are independent and are used to decompose the problem into subproblems along the cost dimension. Identified are projects that are mutually exclusive with respect to construction costs (i.e., projects with costs that do not incur the need of any further expenditures and that exclude the upgrading of any other project).

In this analysis, a project is the upgrading of path-links that allows the establishment of a better load category for an entire path and the realization of benefits. Projects leading to only unrealizable benefits are discarded, so the final list contains only an implicit enumeration of all projects that have realizable benefits. Unrealized benefits of a complete project are not weighted in this implicit enumeration procedure.

MODEL DEVELOPMENT

Road Network Representation

Three characteristics of the road network are of interest: (a) the nature of the network, (b) the load-carrying type of each arc (road), and (c) the length of each arc. This information may be represented in a matrix P of size $N \times N$, where N denotes the number of nodes in the network. The element a_{ij} in cell (i,j) of matrix P is zero, wherever nodes i and j in the network are not directly connected. A nonzero entry occurs only where nodes i and j are connected by a direct arc. For a pair of nodes i and j that are directly connected, the element a_{ij} is represented by a pair of numbers, the first number k giving the weight-carrying type (e.g., 9 ton), and the second number l giving the length of the arc (i,j) (Figure 1).

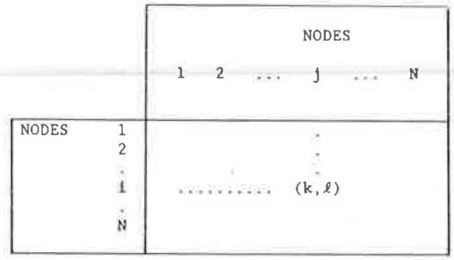


FIGURE 1 Adjacent arc matrix representing the network; a nonzero entry appears only where nodes *i* and *j* are connected by a direct arc.

Customer's Route Demand Matrix

Initial concentration is on a single commodity served by the network, and the demand matrix for this commodity is defined while the index for the commodity is suppressed. The final purpose is, of course, to compute the total net benefits realizable for all commodities from the upgrading of the network. When the scheme for computing the benefit for one commodity has been laid out, the benefit for all commodities can be computed easily by summing the benefits for all individual commodities.

Let (i_s, j_s) denote the pair of source and sink nodes for customer s, where the index for the commodity is suppressed. Let $d(i_s, j_s)$ denote the annual demand in tons for customer s from source node i_s to sink node j_s. The demand $d(i_s, j_s)$ and routing information for each customer s is given in Table 1.

TABLE 1 MATRIX OF SOURCE-SINK PAIRS, ROUTING AND DEMAND DATA

			List of All Arcs						
s	i_s	j_s	E1	E2	E3	F1	F2	G9	$d(i_s, j_s)$
1	3	7	1	1				1	20,000
2	3	7		1	1	1		1	5,000
3	1	8				1	1	1	32,000
.	.	.							.
.	.	.							.
.	.	.							.

The first column of the table, labeled s, indicates the customer number and the next two columns, labeled i_s and j_s, indicate the source and sink node pairs for each customer. The middle section includes the routing information; each row represents a route from a source node to a sink node. Each arc is coded according to existing weight restrictions (e.g., E stands for 9-ton roads operated as 10-ton roads in the three winter months). A 1 under an arc implies that this arc is involved in the route from i to j, and no entry implies that the arc is not involved. More than one route may be listed for each source-sink pair by assigning a separate line to each route.

For each source-sink pair, the last column indicates the annual demand in tons from node i_s to node j_s. As upgrading of the arcs proceeds, it is likely that some customers may change from their current routes to different routes. To provide for this possibility, all routes that can potentially become optimal routes are listed a priori.

Route Capacity

The special feature of the transportation network is that the weight limit on a route is determined by the minimum value of the weight limits of the arc involved in that route. Benefits from upgrading weight limits of various arcs are therefore realized only if these improvements lead to the raising of the minimum value of the weight limits on complete routes. Two maintenance policy alternatives are considered for the route capacity of the network: (a) Arc(i,j) is upgraded from its current type k_{ij} to a new type $\hat{k}_{ij}$, where $\hat{k} > k_{ij}$. (b) Arc(i,j) of current type k_{ij} is used without improvement for loads of type $\hat{k}_{ij}$. This would lead to a reduced expected life and an increase in the maintenance costs. Further, let

$c(i,j,k,\hat{k})$ = present worth of the sum of the initial costs for upgrading arc(i,j) from load type k to load type $\hat{k}$, and the maintenance costs over a planning horizon of T years and

$e(i,j,k,\hat{k})$ = present worth of the increased maintenance costs incurred over a planning horizon of T years when arc(i,j) of load type k is used for load type $\hat{k}$, where $\hat{k} > k$.

For every arc with current load type k, a decision has to be made: should it be improved and, if so, to what new category, or should it be used for higher loads without improvement?

Decision Variables and Mathematical Formulation

Let

$$X(i,j,k,\hat{k}) = \begin{cases} 1, \text{ if arc}(i,j) \text{ is upgraded from type } k \text{ to } \hat{k} \\ 0, \text{ otherwise} \end{cases}$$

$$Y(i,j,k,\hat{k}) = \begin{cases} 1, \text{ if arc}(i,j) \text{ of type } k \text{ is used for loads of type } \hat{k} \\ 0, \text{ otherwise} \end{cases}$$

Then, for every arc(i,j), find $X(i,j,k,\hat{k})$, $Y(i,j,k,\hat{k})$ that maximize the total net benefit Z summed over all s customers and all p commodities:

$$\max Z, X, Y = \sum_{p,s} b_p\,(i_s,j_s,m,n) \tag{1}$$

where the term $b_p\,(i_s,j_s,m,n)$ denotes the net benefit realized for customers of product p as a result of raising the minimum weight limit from m to n on the route from i_s to j_s.

This optimization is subject to the following constraints:

$$\sum_{(i,j)} \left[\sum_{\hat{k}>k} c(i,j,k,\hat{k})\, X(i,j,k,\hat{k}) + \sum_{\hat{k}>k} e(i,j,k,\hat{k})\, Y(i,j,k,\hat{k}) \right] \le W \tag{2}$$

where W is the present worth of the total available budget over the planning horizon of T years and

$$\sum_{\hat{k}>k} [X(i,j,k,\hat{k}) + Y(i,j,k,\hat{k})] \le 1 \text{ for every } (i,j) \tag{3}$$

In Equation 1, m and n define the present and the new load limits for route (i_s,j_s) and these are equal to the minimum values of k and $\hat{k}$, respectively, for the arcs (i_s,j_s) involved in the route (i,j). Therefore, a benefit is realized only when the minimum load limit on a route is raised from m to n.

This problem does not appear to be amenable to an obvious dynamic programming formulation (i.e., one based on Bellman's principle of optimality). According to that principle, if a specific amount of a resource is allocated to a given activity, say activity i, there is a chance of obtaining an overall optimal return only if the remaining amount of the available resource is allocated in an optimal fashion among the remaining activities. The principle does not hold in this case because the set of transportation projects that is the optimal solution for a large budget does not necessarily contain (as a subset) the optimal solution project set of a smaller budget. Because the principle of optimality cannot be used to eliminate a feasible solution of the problem, a branch-and-bound or other programming technique or an enumerative approach may be used. Branch-and-bound has not been considered because the network is of a size for which the enumerative approach is quite adequate. A solution algorithm that is essentially enumerative is developed in the next section.

SOLUTION ALGORITHM

The algorithm for obtaining optimal highway development (e.g., upgrading) plans at any available budget W is based on complete enumeration, a reasonable strategy when the size of the problem is not too large. The algorithm follows three basic steps:

• Step 1: Generate a set U of all feasible combinations of elemental projects, coded by highway arc. Arrange these projects in a monotonic increasing order based on their cost.

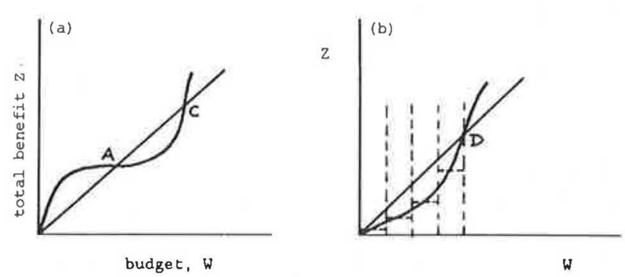


FIGURE 2 Total optimal benefit as a function of available budget: a, beginning with B/C > 1; b, beginning with B/C < 1.

- Step 2: From set U, generate a set V that indicates all of the feasible breakpoints on the budget axis (Figure 2).
- Step 3: For any given budget W, select the set of projects that maximizes total net benefit. Repeat for all breakpoints on the budget axis.

The projects are initially ordered on the basis of cost (Step 1) merely to facilitate the subsequent search for feasible budget breakpoints (Step 2). Project selection then proceeds according to any acceptable criterion such as net benefit or benefit-to-cost ratio (B/C). In this analysis an elemental project is defined as the upgrading of a route from node i to node j that allows the establishment of a better load category for the entire route (i,j). Further, the set U of all feasible project combinations includes upgrading combinations that lead to the same final outcome but are accomplished in a different sequence. For instance, a 9-ton road may be upgraded to 10 tons directly; alternatively (and this would be considered a different project in U), the 9-ton road may be partly improved at first, to 10 tons for 10 months. To be sure, the cost of upgrading a highway in steps is higher than making the complete improvement all at once.

After a project has been selected for completion, the cost of all arcs belonging to that project is set equal to zero and the costs of all remaining projects are updated. For projects that include arcs that are common to those of the selected project, the cost decreases; for all others, the cost remains the same.

The nature of the relationship between the total optimal benefit Z and the available budget W is shown in Figure 1. In general, the set U initially may contain one or more small projects the completion (upgrading) of which leads to immediate completion (upgrading) of one or more complete routes. If such projects are present in U, the curve of Figure 2a begins with a B/C ratio greater than one. If, on the other hand, no such project exists in U initially, the rate of accumulation of the total benefit Z is slow and the curve begins below the break-even (B/C = 1) line as Figure 2b indicates. As more arcs are completed, the benefit accumulation rate accelerates and the curve of Figure 2b may again cross the break-even line as it enters a range where B/C > 1 at some stage (Point D). Toward the end, when most important routes in the network have been upgraded, the rate of increase of Z slows down again.

It should be noted that, when the budget is overly restricted or the highway network is well developed, the B/C curve of Figure 2 may end as convex (i.e., reaching the break-even line from below rather than from above), Points C or D in Figure 2 may then never be reached. Further, the continuous curve of Figure 2 should, more accurately, be discrete reflecting the discrete nature of the optimal benefit increments (see the dashed lines in Figure 2b).

CASE STUDY IN NORTHEASTERN MINNESOTA

Case Description

The objective of the case study is to analyze the economic viability of upgrading the spring weight restrictions on the state highways of northeastern Minnesota. In particular, the case study is focused on evaluating upgrading the network on the basis of realized net benefits from the paper and waferboard product industries of that region. Benefits would accrue if network upgrading reduced transportation costs and, thus, made the final production cost of these forest products more competitive in the nation's markets. These industries could, then, increase the production capacity of their plants and, in time, their market share in the national and international markets.

Although transportation cost is an important factor in the final cost of voluminous forest products, organized cost and shipment data do not exist or are incomplete. In particular, the difficulties associated with the collection of reliable data and data confidentiality are often cited (*18*, *19*) as the two major reasons for the lack of complete data. To obtain a more complete data base on paper and waferboard product shipments, a survey was conducted in northeastern Minnesota in 1985. The survey sought information on shipment origins and destinations, cost structure, tonnage, modal split, shipment value, trip duration, and the like for the nine leading pulpwood mills in the area. The paper and waferboard producers belonged to the following companies: Potlatch, Blandin, Northwood Panelboard, Boise Cascade, Superwood, Conwed, Diamond International, and Great Lakes Forest Products. A summary of relevant data from these producers is given in Tables 2 and 3.

In addition to the information summarized in Tables 2 and 3, the responses to the survey indicated that transportation cost is an important component of the final price of paper and waferboard, especially for shipments outside Minnesota. To be sure, each company has established its own transport policy that may not necessarily include the transport cost explicitly in the final product price. Further, not all companies collect information on transport cost components (such as travel time and loading cost) in a uniform manner, and a substantial portion of it is based on estimates. In general, the paper market is relatively more stable than the waferboard market; it employs more trains over longer distances, and procurement planning is more long term. Waferboard planning is based on a shorter horizon and involves shorter haul and heavier use of trucks whose drivers often determine their own routes. No surveyed company disclosed product demand data at the customer, town, or city level. As a result all demand data are at the state level.

The data base was expanded with data related to the principal highways the forest industries use in northeastern

TABLE 2 ACTIVE PULPWOOD MILLS AND WAFERBOARD PLANTS IN NORTHEASTERN MINNESOTA BY LOCATION AND CAPACITY, 1982 (*18*)

Company	Location	Capacity[a]
Pulpwood mills		
Producer X	Grand Rapids	300
Producer Y	International Falls	920
Producer Z	Cloquet	475
Producer T	Bemidji	100
Producer U	Duluth	350
Producer V	Cloquet	50
Waferboard plants		
Producer A	Grand Rapids	270,000
Producer B	Bemidji	160,000
Producer C	Bemidji	150,000
Producer D	Cook	150,000

[a]Capacity for pulpwood mills is tons/24 hr; for waferboard plants, capacity is estimated tons per year.

TABLE 3 DATA SUMMARY OF FOREST PRODUCT PRODUCERS IN NORTHEASTERN MINNESOTA

Company	Maximum Distance Between Plant and Market (mi)	Maximum Distance Between Plant and Market by Truck (mi)	Transportation Cost ($/mi)		Shipment Size (short tons)		
			Truck	Rail	Flat-Bed Truck	Irregular Common Carrier	Rail
1	1,800	U	1.2	2.5	23	23	na
2	2,100	U	1.1–1.4	2.2–5.5	na	na	51
3	48 states	800	1.2	3.5	23	na	75
4	48 states	800	1.2	3.5	23	na	75
5	48 states	U	1.1–1.3	na	na	23	na
6	48 states	U	1.1–1.2	na	23	18	na
7	700	700	NA	1.0	na	22	62

NOTE: For purposes of confidentiality, not all producers are listed. U = unlimited; depends on market conditions and order size. na = not applicable; mode not used. NA = not available.

Minnesota. These data, provided by the Minnesota Department of Transportation (MnDOT), were used to develop the layout of the relevant highway network, shown in Figure 3. The MnDOT classifies these highways in three load categories:

- E category: 9-ton roads operated at 10 tons in the 3 winter months,
- F category: 9-ton roads operated at 10 tons for 10 months, and
- G category: 10-ton roads year-round.

This information was used to segment the principal highways of northeastern Minnesota into links by load category and estimated remaining life (Table 4).

After the relevant links had been identified and classified, the algorithm was implemented to analyze these highways with the help of a personal computer using Pascal. The computer code accepts the arc length and remaining life of highways and the number of truckloads between origins and destinations as inputs. The output is a priority list of the available projects subject to a budget constraint. The results are based on the assumption that the realizable project benefit per truckload is approximately 3 short tons [i.e., the difference between the currently allowed 73,820-lb gross vehicle weight (GVW) and the desirable 80,000-lb GVW]. No effects were considered that relate to possible truck detouring or plant closing because of road deterioration.

Case Study Results and Discussion

Before the results of this case study are discussed, a few comments are in order regarding the relevance of this case to

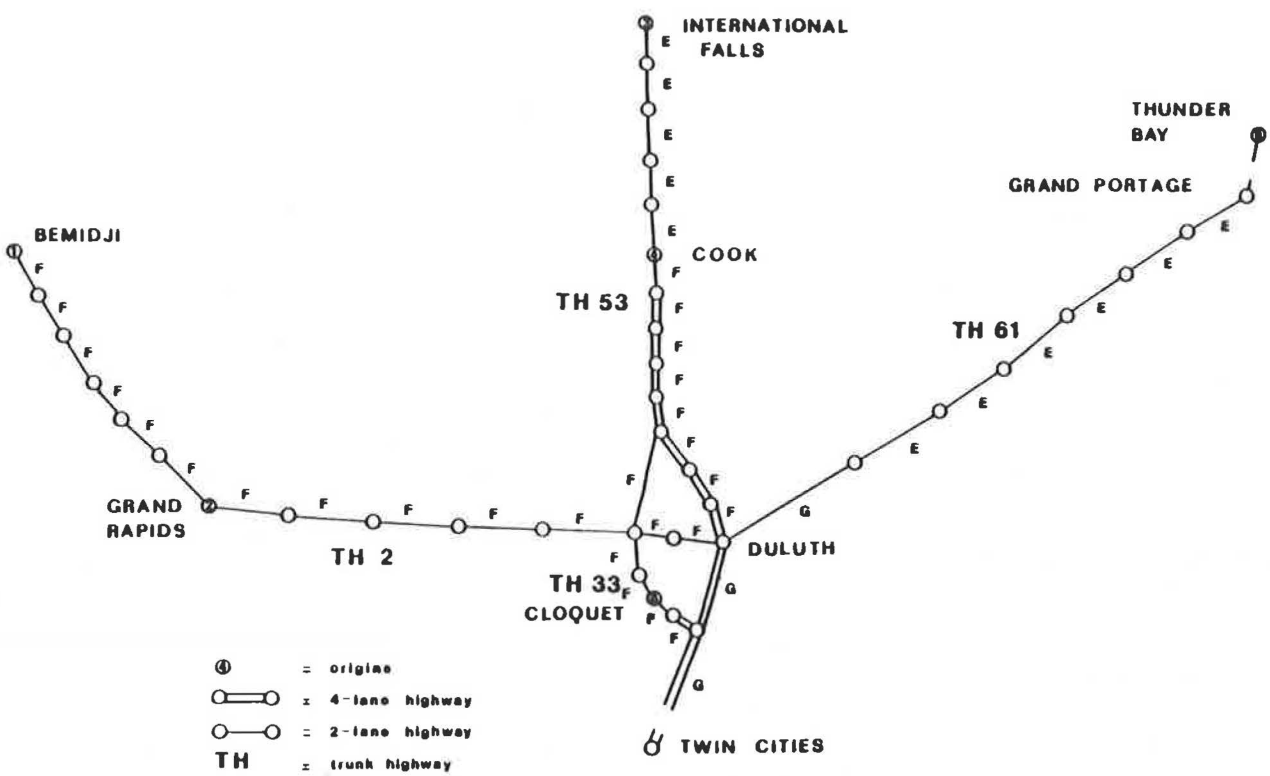


FIGURE 3 Principal highways of northeastern Minnesota used by forest industries.

TABLE 4 TRUNK HIGHWAYS OF NORTHEASTERN MINNESOTA USED BY FOREST INDUSTRIES

Trunk Highway (TH)	Node	2 Lanes		4 Lanes		Category
		Mileage	Remaining Life (yr)	Mileage	Remaining Life (yr)	
33	I-35	2.65	4	2.65	4	F
		1.03	25			F
	Cloquet	3.75	9			F
		4.10	24			F
	TH-2	8.25	8			F
	TH-53					
2	Duluth	7.01	20			F
		6.77	35			F
	TH-194	1.94	12			F
	TH-33	12.62	20			F
		25.65	17			F
		7.38	23			F
		12.24	5			F
		2.85	9			F
	Grand Rapids	9.9	18			F
		0.35	9			F
		12.35	38			F
		6.95	14			F
		27.35	29			F
		21.8	24			F
	Bemidji					
53	Duluth	7.0	6	7.25	6	F
		8.64	10	9.94	10	F
		2.91	21	1.36	21	F
	TR-33	1.67	21	12.03	21	F
		16.97	7	8.54	7	F
		16.03	4	16.09	4	F
		11.85	18	9.86	18	F
		22.21	19			F
	Cook	2.00	38			E
		29.76	23			E
		17.26	17			E
		18.07	11			E
		3.25	9			E
	International Falls					
61	Duluth	12.00				G
	Two Harbors	3.55	20			E
		17.77	24			E
		27.43	22			E
		38.41	15			E
		16.14	11			E
		17.70	8			E
	U.S. border					

typical project selection and priority-setting problems. More specifically, upgrading highway weight limits has been a major issue in the state of Minnesota and the choice of the particular topic is, therefore, timely. The upgrading issue is particularly relevant in the north where road condition requires extensive improvement.

The issue is also relevant in that part of the state for two additional reasons. First, the timber industry is a major user of the roads; that industry carries heavy loads over long distances and is incurring a substantial competitive disadvantage by having to operate trucks below capacity. Therefore the industry has been vocal in its requests for road upgrading. Tourism is the second major user of the roads in the north and could benefit from improved road quality. In particular, previous findings indicate that tourist-related services stand to gain substantially when access is improved. These findings were recently confirmed for Minnesota (2), where it was found that only in nonmetropolitan counties that have a strong tourist base do highway improvements have a significant long-term beneficial effect on employment.

Although the importance of both the timber industry and tourism to the economy of northeastern Minnesota is recognized, time limitations allowed this method to be implemented with timber movements only. Therefore the determination of benefits that would result from upgrading is conservative because it only includes timber-related benefits.

Of all candidate upgrading projects considered, the following were selected in order of priority, based on the selection algorithm (the selected projects are shown on the Minnesota map of Figure 4) and the estimated benefits that would result for the timber industry:

1. TH-33 from I-35 to Cloquet: upgrade to 10-ton road year-round.
2. TH-33 from Cloquet to TH-2 and TH-2 from TH-33 to Grand Rapids: to 10-ton road year-round.

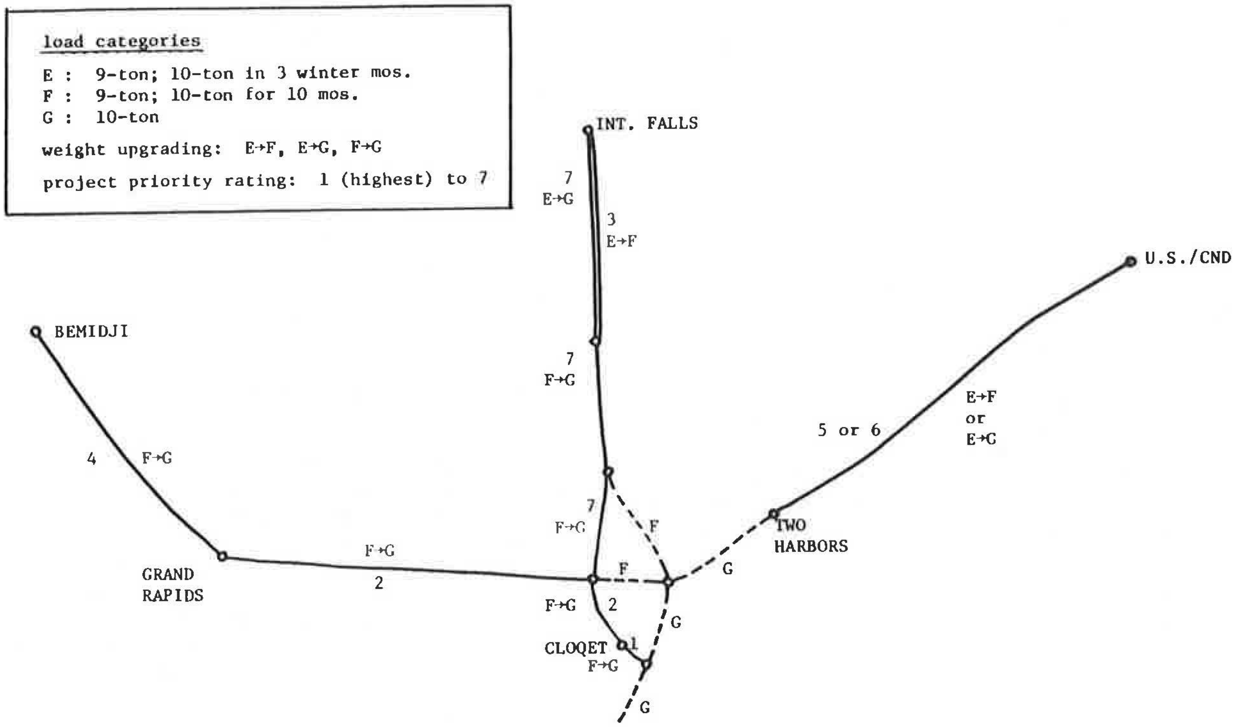


FIGURE 4 Projects in northeastern Minnesota in priority order.

3. TH-53 from Cook to International Falls: to 10-ton road for 10 months.
4. TH-2 from Grand Rapids to Bemidji: to 10-ton road year-round.
5. TH-61 from Two Harbors to U.S. border: to 10-ton road for 10 months or year-round.
6. TH-33 from TH-2 to TH-53 and TH-53 from TH-33 to International Falls: to 10-ton roads year-round.

It was noted that, when selections 1 and 2 from this set have been made, the remaining selections indicate a cumulative B/C ratio that is less than 0.2 and may, thus, not appear attractive at this stage. Indeed, only the segment of Trunk Highway 33 (TH-33) connecting Interstate 35 (I-35) with Cloquet (Figure 4) has a B/C ratio greater than 1 if only timber-related travel is considered.

This finding is not surprising and does not indicate lack of relevance of the new method. The low cumulative B/C is partly the result of considering the benefits accruing to only one customer, the forest industry. When the benefits accruing to the additional economic sectors that stand to benefit from improved access (such as the service sector in relation to tourism) are considered, the B/Cs of these projects are expected to improve. It was noted that the project priority-setting algorithm was effective in reducing a quite large number of possible project combinations to a priority list of manageable size. Having considered the estimated benefits for only one industry and the upgrading costs, the priority-setting algorithm conclusively indicated the desired order in which the projects should be undertaken. Priority setting could certainly be extended to consider expected benefits to additional industries. This analysis does not consider the opportunity cost of not tending to deteriorating highways in a timely fashion. For instance, roads of low quality are likely to result in truck detours, when an alternative path is available, and higher transportation cost. When the cost crosses a certain threshold, which the industry considers unacceptable, the industry may relocate; similarly, new industry may not be attracted. Further, the analysis does not consider any rerouting that may take place after partial upgrading of the network. However, the centralized nature of the northeastern Minnesota network substantially reduces the possibility for such rerouting.

It should be noted that the MnDOT has recently decreased weight restrictions on TH-2 on the basis of highway engineering criteria (deflection tests) and is considering upgrading TH-33 from I-35 to Cloquet. These decisions, made independently of this analysis, are in substantial agreement with the present results.

SUMMARY

A heuristic framework was developed for the selection and priority ranking of highway weight upgrading projects. The method can help the decision maker identify the most worthwhile projects in terms of benefits to highway users and upgrading costs over the planning horizon. The analysis evaluates all feasible project combinations. In particular, it considers all individual highway arcs of each project in every order and all combinations of intermediate upgrading possibilities. A special constraint of the problem dictates that a benefit for a path is realized only when the minimum load limit along the whole path is raised.

Without loss of generality, the method was applied to the northeastern Minnesota network to evaluate all possible

upgrading project combinations relative to a major highway user, the forest industry. Following the evaluation, the long list of possible project combinations led to the identification of a small set of projects that were ranked in priority order for implementation. It is encouraging to note that, even though the example application was limited to one user, the results of the priority ranking are in substantial agreement with the upgrading decisions that the MnDOT made independently of this analysis.

Although the algorithm leads to a conclusive priority listing of the best project combinations selected from an all-inclusive list of feasible projects, it must be used for each major highway user in order to reflect the benefits that would accrue to all users. The algorithm was implemented in a case study that was limited to only one industry, but its extension to additional industries is straightforward because it has been designed to be used in the general case of the highway user. Ongoing research seeks to include the time element in the analysis. For instance, it is desirable to identify the time at which each of the reviewed projects may become attractive subject to a planning horizon and annual budget restrictions.

ACKNOWLEDGMENT

The work described here is the result of research funded by the Minnesota Department of Transportation. The comments made by three anonymous referees and the members of the Committee on Transportation Programming, Planning and Systems Evaluation of the Transportation Research Board are gratefully acknowledged.

REFERENCES

1. Y. J. Stephanedes and D. M. Eagle. Time-Series Analysis of Interactions Between Transportation and Manufacturing and Retail Employment. In *Transportation Research Record 1074*, TRB, National Research Council, Washington, D.C., 1986, pp. 16–24.
2. Y. J. Stephanedes. *Influence of Transportation on Economic Development.* Final Report MnDOT-59757M-62179-III. Department of Civil Engineering, University of Minnesota, Minneapolis, 1988.
3. *NCHRP Synthesis of Highway Practice 84: Evaluation Criteria and Priority Setting for State Highway Programs.* TRB, National Research Council, Washington, D.C., 1981.
4. *Transportation Research Circular 213: Bibliography on Project Evaluation and Priority Programming Criteria.* TRB, National Research Council, Washington, D.C., Jan. 1980.
5. *NCHRP Synthesis of Highway Practice 48: Priority Programming and Project Selection.* TRB, National Research Council, Washington, D.C., 1978.
6. B. G. Bisson, J. R. Brander, and J. D. Innes. Highway Cost Allocation Methodology for Pavement Rehabilitation and Capacity-Related Costs Occasioned by an Increment in Heavy Truck Traffic. In *Transportation Research Record 1038*, TRB, National Research Council, Washington, D.C., 1985, pp. 10–16.
7. M. J. Markow and T. K. Wong. Life-Cycle Pavement Cost Allocation. In *Transportation Research Record 900*, TRB, National Research Council, Washington, D.C., 1983, pp. 1–10.
8. *Final Report on the Federal Highway Cost Allocation Study.* FHWA, U.S. Department of Transportation, 1982.
9. *Highway Maintenance.* Program Evaluation Division, Office of the Legislative Auditor, State of Minnesota, St. Paul, Minn., 1985.
10. J. A. Gomez-Ibanez and M. M. O'Keeffe. *The Benefit from Improved Investment Rules: A Case Study of the Interstate Highway System.* Final Report DOT/OST/P-34/86-030. Harvard University, Cambridge, Mass., 1985.
11. G. Bergendahl. A Combined Linear and Dynamic Programming Model for Interdependent Road Investment Planning. *Transportation Research*, Vol. 3, No. 2, 1969, pp. 211–228.
12. P. A. Steenbrink. Transport Network Optimization in the Dutch Integral Transportation Study. *Transportation Research*, Vol. 8, No. 1, pp. 11–26.
13. J. E. Gruver, F. P. Patron, and R. D. Juster. Highway Investment Analysis Package. In *Transportation Research Record 599*, TRB, National Research Council, Washington, D.C., 1976, pp. 13–18.
14. H. Schnuerer. Optimization of Road Investments Based on Accessibility Criteria Using a Dynamic Programming Model. *Transportation Planning and Technology*, Vol. 9, 1984, pp. 237–246.
15. H. Muller-Merbach. Heuristic Procedures for Solving Combinatorial Optimization Problems in Transportation. *Transportation Research*, Vol. 10, No. 6, 1976, pp. 377–378.
16. H. Crowder, E. L. Johnson, and M. Padberg. Solving Large-Scale Zero-One Linear Programming Problems. *Operations Research*, Vol. 31, No. 5, 1983, pp. 803–841.
17. T. L. Magnanti and R. T. Wong. Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, Vol. 18, No. 1, 1984, pp. 1–55.
18. C. Eldridge and J. Fruin. *The Transportation of Minnesota Forest Products.* University of Minnesota, St. Paul, 1984.
19. *Regional Goods Movement Study.* Arrowhead Regional Development Commission, Duluth, Minn., 1985.

Publication of this paper sponsored by Committee on Transportation Programming, Planning and Systems Evaluation.