# Decision-Making Procedure for Assessing Performance Measures of Freeway Operations

Auttawit Upayokin, Stephen P. Mattingly, and Sergio A. Lugo-Serrato

Performance measurement is a method to quantify project success based on project and policy outputs and outcomes assessment. In the United States, transportation-related performance measures have become popular since authorization of the Intermodal Surface Transportation Efficiency Act in 1991. Performance measures have been applied in many areas, including freeway operations. This paper improves on earlier techniques and presents an innovative approach for using a multicriteria decision framework to select freeway operational performance measures. The proposed selection performance measure methodology includes establishing a statement of purpose, identifying alternatives, establishing criteria, and using a screening approach to eliminate unsatisfactory performance measures by means of qualitative and quantitative criteria. Once the screening approach is used to reduce the set of candidate alternatives to a set of feasible alternatives, the methodology uses a multicriteria decision model, such as simple additive weighting and ELECTRE III, to combine the commensurate criteria and generate a rank order of the acceptable alternatives for final decision making. A sample application of the proposed methodology is presented.

Selecting freeway performance measures is a complex and difficult process that requires considering the different aspects influencing freeway operations. Furthermore, construction funds, labor, and materials are becoming increasingly scarce, which reduces the possibility of decreasing congestion problems by increasing capacity (1). Thus, traffic operations management plays an important role in alleviating traffic congestion and in improving safety and mobility on existing freeway systems. Operations management strategies used for freeway systems include traffic incident detection, traveler information systems, managed lanes, and ramp management. To evaluate freeway performance before and after any of these operational strategies are implemented, measures of effectiveness (MOEs), such as density, speed, volume, and travel time index, are used (2, 3). The quality of these performance measures depends on the data collection techniques and equipment used. Transportation professionals must determine whether equipment and data collection techniques should be provided in specific locations or exclusively when incidents occur; some equipment can provide performance measure data without additional investment (e.g., existing loop detectors provide spot speeds).

Department of Civil and Environmental Engineering, University of Texas at Arlington, P.O. Box 19308, Arlington, TX 76019-0308. Corresponding author: Stephen P. Mattingly, mattingly@uta.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2046, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 11–19.

DOI: 10.3141/2046-02

Recently, NCHRP Synthesis of Highway Practice 311 developed a scoring approach to assess the strengths and weaknesses of performance measures. Qualitative criteria performance measures include clarity and simplicity, descriptive and predictive ability, analysis capacity, accuracy and precision, and flexibility (3). These criteria may do a good job of assessing performance measure quality in a general sense; however, they do not provide a strong approach for identifying preferred performance measures based on specific local characteristics, such as the reliability and accuracy of different data collection strategies for the same performance measure. In addition, when a particular performance measure is chosen, there will most likely be either costs associated with the collection of the data necessary for generating the performance measure in the first place or ongoing operational costs associated with tracking and maintaining the data themselves. The decision-making criteria should include commensurate quantitative criteria because decision makers are unlikely to arrive at the preferred alternative when using only general assessments of qualitative criteria.

This paper provides a decision-making framework for assisting transportation professionals in their selection of alternative qualitative freeway performance measures, such as understanding and measurability, and quantitative freeway performance measures, such as cost and accuracy. This research considers the scoring of qualitative criteria in NCHRP Synthesis 311 (3) as a potential starting point for selecting among alternative performance measures from which decision makers can decide whether to consider or reject a performance measure based on its score; decision makers can establish minimum performance thresholds for qualitative criteria, and the performance measures that fail to meet or exceed all of these thresholds will be excluded from further consideration. Remaining candidate performance measures must still be evaluated for local characteristics using quantitative criteria. Decision makers may establish similar performance thresholds for quantitative criteria and use them for screening as well. The authors propose using a multicriteria decision model (MCDM), such as simple additive weighting (SAW) or ELECTRE III (4), after all the screening is complete to combine quantitative criteria for evaluating the performance measures. An MCDM ranks the alternatives for the decision makers to make their final selection. The proposed approach improves on existing techniques by adding a screening stage and integrating quantitative criteria that can handle local characteristics. These improvements make the approach viable for local decision makers who must choose freeway operational performance measures under budget and personnel constrained conditions. The rest of this paper discusses the proposed methodology in detail and presents an example of its application.

#### **DEVELOPING PERFORMANCE MEASURES**

Before a discussion is begun on how performance measures are assessed, the way performance measures are developed must be considered. One possible use of performance measures is in the assessment of freeway operational strategies. Planners can use a top-down methodology in which the types of performance measures may be established at the highest level to assess the program goals, which may not be easily measurable. Then at the next lower level, either output or outcome measures may be used to evaluate the program objectives. At the next lower level, output measures are commonly used to assess the immediate impacts of policies or projects. At the lowest level, input measures are generally used to assess program resources. The use of program targets may be integrated into this process to provide ongoing monitoring and assistance with future improvement decisions. Figure 1 is an example of a transit system improvement that shows the impacts at each level.

The top-down methodology can be used in multilevel operations to assess freeway performance as shown in Figure 2. The application of performance measurement in freeway operations can incorporate multiple scales or levels based on the number of agencies using performance measures. At the top level, systemwide performance measures are used to assess a global view of operations. In the next lower level, called the interagency level, many agencies will share resources to improve their operational programs, such as incident management and air quality. The third level, daily operations, focuses on Transportation Management Center (TMC) operations, such as lane shifts, dynamic messages, signal timing, and ramp metering. TMC operators may use these measures to assess their programs and strategies. At the bottom level, measures are used to assess equipment or discrete elements of the transportation system, such as equipment reliability. Because performance measures can be used in multiple scales, good performance measures should be able to assess the freeway performance in multiple scales also. For example, performance measures used to assess at the systemwide level should also apply at the interagency level (5).

Because performance measurement has been used in many fields, there is no single methodology or exact rule for selecting specific measures. In addition, criteria for selecting appropriate performance measures should be decided by the people who are involved in the performance measurement program, such as those who collect and use data or experts who understand the strengths and limitations of each performance measure. Good performance measures, in general, should focus on the goals and objectives of the program

whose performance is to be assessed. They should be simple, easy to understand for everyone, able to respond to changes in the system, inexpensive to obtain, organizationally acceptable, credible, timely, comparable, compatible, customer focused, consistent, measurable, available, balanced, valuable, and practical (5–7). The following section provides a framework for developing appropriate performance measures.

### PERFORMANCE MEASURE CREATION AND SELECTION METHODOLOGY

The five steps developed in this paper focus on the roles of decision makers and coordinators in creating and screening performance measures. Decision makers are defined as the people responsible for using performance measures for evaluating freeway operational strategies, and coordinators are defined as people who assist decision makers by guiding the process and conducting any necessary analysis during the performance measure selection process.

### Step 1. Establish the Decision Statement

Specifying the proper decision statement is a crucial step for decision makers because it can determine whether the solution meets the desired goal. In addition, the established decision statement will lead to the solutions' being simple versus complex or broad versus narrow. For example, the decision statement "select a strategy to reduce traffic congestion on a freeway" will provide a solution of only the one best strategy. However, if the decision statement is changed to "select strategies to reduce traffic congestion on a freeway," it may lead to the consideration of various strategies that minimize traffic congestion on a freeway. The roles of decision makers and coordinators involved in this step are as follows:

- Decision makers have to clarify the objective by identifying a problem and establishing a decision statement to scope the problem's boundary and candidate solutions. A problem can cause various challenges, for example, traffic congestion may lead to environmental externalities, such as air pollution. Thus, decision makers should focus on the main cause of the problem rather than the outcomes.
- Coordinators should ensure that the results of the discussion lead to the problem's main cause. They should provide useful information, which includes traffic condition data, travel behavior, and on-road activities.

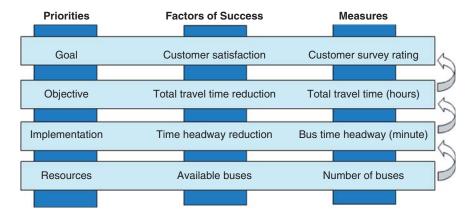


FIGURE 1 Multilevel structure for performance measures.

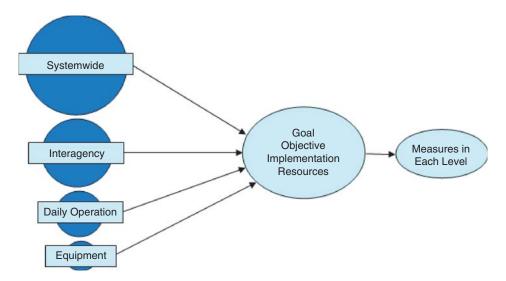


FIGURE 2 Multilevel operation approach.

### Step 2. Identify the Set of Alternatives or Solutions

Once a decision statement is identified, decision makers must clearly understand the program goals and performance measures. They must establish the possible alternatives or solutions based on the decision statement described in Step 1. The roles of decision makers and coordinators involved in this step are as follows:

- Decision makers must establish the candidate alternatives or solutions and
- Coordinators should provide any additional information as needed.

The useful information will enhance decision makers' viewpoints in the selection of candidate alternatives or solutions. To accomplish this, coordinators may conduct a survey for decision makers that may include the following questions:

- What are the objectives of the program?
- What are the current operational performance measures used in the program?
- What operational performance measures do you expect to be applied in the future?

The purpose of providing the program objectives is to make sure that decision makers consider these objectives when generating candidate alternatives.

### Step 3. Establish the Set of Criteria to Assess Performance Measures Based on Equipment and Data Collection Techniques Used on Freeway Systems

Once the set of candidate alternatives is generated in Step 2, the set of constraints for identifying the feasible alternatives should be established. The set of constraints may include qualitative and quantitative criteria. The qualitative criteria will be used to determine the possibility of assessing the performance measures, and they

may include measurability, comprehension, and availability. The quantitative criteria may include budget constraints or limitations of equipment and data collection techniques. The following are the roles of the decision makers and coordinators involved in this step:

- Decision makers have to identify the "set of constraints" for assessing the quality of the freeway performance measures in the next step. They should indicate the critical criteria.
- Coordinators should provide any additional information, as needed, especially information that enhances the decision makers' criteria selection. The information should include the limitations of the candidate alternatives provided in the previous step. The possible questions for decision makers include the following:
  - What are the current performance measures used in the program at the systemwide, interagency, daily operation, or equipment level?
    - What factors affect the use of those performance measures?

In Steps 2 and 3, various techniques, such as brainstorming, nominal group technique (NGT), surveys, or the Delphi method, may be used to generate candidate alternatives and their criteria. Coordinators need to select the proper technique for a particular situation because each technique has its own advantages and limitations. For example, according to Ababutain (8), the objective of brainstorming is to generate all possible ideas to enhance the possibility of reaching ideal solutions. Thus, the final results may generate an unlimited number of solutions; however, the limitation of brainstorming occurs when some members of the group have strong opinions, leading other members to quickly reach an agreement without a complete discussion. Thus, the results may not include other potentially better solutions. Unlike the brainstorming technique, NGT uses a questionnaire survey to allow decision makers to communicate in writing, which can avoid preliminary arguments. However, surveys that allow faceto-face and phone interviews may create conforming influences and decrease the possibility of generating ideas freely. Because selection of freeway operational strategies uses experts who are directly involved with the freeway management system, the researchers recommend the Delphi method. According to Dalkey and Helmer (9), coordinators will select respondents from a group of experts that will be asked intensive questions with controlled opinion feedback. Disagreements among experts will develop successive iterations until various opinions yield to a widely acceptable view. The process ends at that point. The success of this method depends on experts' knowledge, experience, and viewpoints, which can reflect the true value of whatever they judge. For example, constructing a new freeway can provide both advantages and disadvantages depending on the experts' viewpoints. A new freeway can reduce traffic congestion; however, it can induce new vehicles to use it and increase on-road emissions.

## Step 4. Screening the Set of Alternatives or Solutions in Step 2

Quality is better than quantity. More performance measures do not mean that they will provide a better assessment of the program. Thus, Step 4 provides an approach for screening candidate alternatives and identifying feasible alternatives, which can be used to assess freeway operations; then Step 5 uses MCDM models to rank the feasible alternatives. Within Step 4, the processes include grouping the performance measures, defining the direct or proxy performance measures, setting constraints, and eliminating performance measures based on minimal assessment levels established by decision makers.

## Step 4.1. Grouping the Alternatives That Convey the Same Meaning

To avoid using redundant performance measures, performance measures that convey the same meaning must be grouped. For example, when planners consider human health impacts due to traffic congestion, the air quality index (AQI) used in the United States and the air quality health index (AQHI) used in European countries are possible indicators for assessing ambient air quality. However, the indices convey the same meaning; to be practical, decision makers should select either AQI or AQHI, not both. The coordinators must clearly understand the definition of the alternatives obtained through Step 2 before they group them.

### Step 4.2. Defining Direct or Proxy Performance Measures

Ideally, decision makers should consider direct performance measures before resorting to proxy performance measures. Good performance measures should be a direct consequence of activities. For example, if the desired result is minimizing traffic congestion, traffic volume should be a good performance measure (direct consequence). A proxy measure, such as vehicle registration, may sometimes be used in the absence of suitable performance measures due to time, budget constraints, or unavailability of data. Unfortunately, a proxy measure may not provide appropriate results because it relies on strong correlation between the factors. Coordinators have to define the performance measures in Step 4.1 as either direct measures or proxy measures.

### Step 4.3. Setting the Constraints for Screening Set of Alternatives or Solutions in Step 2

Qualitative criteria and quantitative criteria will be used to screen candidate performance measures for feasibility in the next step. Both qualitative and quantitative criteria established in Step 3 may have either a minimum or maximum acceptable value for each criterion, which represents the threshold that the performance measures

must reach. Decision makers have to establish the set of constraints and their thresholds.

### Step 4.4. Eliminating the Alternatives by Aspects

Once the set of constraints is established in Step 4.3, alternatives that do not meet the standard thresholds of the criteria are eliminated. Coordinators have to screen the alternatives with qualitative and quantitative criteria using the thresholds provided in Step 4.3.

The qualitative criteria used in this paper include the following:

- Comprehension. Performance measures should be understandable at any managerial level without defining the terminology.
  - Measurability. Performance measures should be measurable.
  - Availability. Performance measures should be readily available.
- Comparability. Performance measures should be comparable with those of other agencies.
- Importance. Performance measures should be useful for the public.

The quantitative criteria suggested in this paper's example include the following:

- Time, including data aggregation time, data processing time, and updating data frequency time;
- Cost, including capital costs, operational costs, and maintenance costs;
- Accuracy, including data processing accuracy, instrumental accuracy, data aggregation accuracy, and human accuracy; and
- Reliability, including failure of field equipment, communication, and database.

# Example of Screening the Set of Alternatives or Solutions in Step 4

The National Transportation Operation Coalition recommends performance measures for local administrators to be used for internal management, external communications, and comparative measurements as follows: customer satisfaction, extent of congestion—spatial, extent of congestion—temporal, incident duration, recurring and nonrecurring delay, speed, throughput-person, throughput-vehicle, travel-link, travel reliability, and travel trip (10). Assuming that the TMC or third level in Figure 2 will use speed as one of the alternative performance measures for a freeway system, the screening approach outlined in Step 4 should be applied as described below.

## Step 4.1. Grouping the Alternatives That Convey the Same Meaning

Speed can be divided into two groups: spot speed and space mean speed. Spot speed is described as an instantaneous speed measured at a specific location; space mean speed is an average travel speed over a distance. The following equipment and techniques are used for collecting spot and space mean speed data:

- Spot speed—gun radar, loop detector, microwave sensor, video sensor, infrared sensor, and acoustic sensor and
- Space mean speed—test vehicle (floating car) technique, license plate matching technique, video matching technique, ITS probe

vehicle technique, time-lapse photography, and toll tag matching technique.

## Step 4.2. Defining Direct or Proxy Performance Measures

Spot speed and space mean speed are both defined as direct performance measures.

### Step 4.3. Setting the Constraints for Screening Set of Alternatives or Solutions in Step 2

The qualitative constraints in this paper include qualitative criteria such as understanding, measurability, availability, comparability, and usefulness; quantitative constraints are cost, accuracy, reliability, and data processing time. Decision makers will select the performance measures that meet both qualitative and quantitative criteria.

### Step 4.4. Eliminating the Alternatives by Aspects

When qualitative criteria are considered, speed, which can be understood by most people, should be an appropriate performance measure; moreover, it is measurable because data can be collected using a variety of data collection techniques. Finally, it can be compared between agencies and is useful for the general public. When quantitative criteria are considered, it is assumed that all equipment is within budget; decision makers will consider the data collection strategies whose accuracy and reliability are above 90% and 95%, respectively, and for which data processing is less than 15 min. Figure 3 illustrates the results when criteria are screened by minimum performance thresholds. The alternatives that do not meet the specified standards are eliminated until remaining alternatives pass all constraints.

# Step 5. Assessing the Set of Alternatives from Step 4

Once decision makers screen the performance measures in Step 4 to determine the feasible alternatives, multicriteria decision making (MCDM) models, which allow decision makers to evaluate the trade-

offs among the various criteria, are applied. MCDM models require decision makers to establish each criterion's weight. Numerous techniques are used for assigning the criteria weights, such as presumption of equal weights, ranking system, ratio system, basic pairwise analysis, one hundred point system, and swing weight approach (4). Each technique has its own advantages and limitations. For example, "presumption of equal weights" may be first assumed when decision makers are not ready to assign the criteria weights; however, equal weights rarely exist in the final set of weights. In addition, a ranking system is a simple method, requiring less effort from decision makers, but the results from the ranking method are only ordinal, which limits the validity of most mathematical operations related to them. In this paper the technique called the "one hundred point system" is used. In this technique, decision makers are given one hundred points to distribute among the criteria. The points allocated to a criterion directly give its percentage weight. The advantages of this method are its straightforward concept and its ability to provide data that can be evaluated using a ratio scale. However, this strength leads to a concern about the results' validity. The coordinator must address this issue through a series of verification questions that confirm that the final weights match the decision makers' intentions. Specifically, the decision maker must agree with the ratio scale implications of his or her criteria weights. The role of decision makers and coordinators involved in this step are as follows:

- Decision makers have to establish the criteria weights through one of the weighting techniques described above. MCDM models will be used to assess those alternatives in the final decision-making process.
- Coordinators have to provide additional questions to assess the consistency of the weighting technique's results. For example, in the "one hundred point system" approach, the scores that a decision maker provides must be comparable with each other; therefore, the relative weights that are derived from the scores must be presented to the decision maker to verify the results.

#### **MULTICRITERIA DECISION-MAKING MODELS**

The selection of performance measures is usually based on more than one criterion; an MCDM approach allows decision makers to analyze complex decision problems with usually conflictive and opposite points of view. There are two main families of MCDM models: utility function-based models and outranking methods (11).

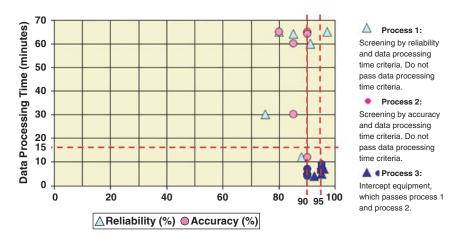


FIGURE 3 Graphic presentations of accuracy, reliability, and data-processing time of data collection equipment and techniques.

The utility function-based models include multiattribute utility theory (12) and simple additive weighting (SAW). Outranking methods include the ELECTRE families (13-15). This paper does not contain a discussion of all multicriteria models in detail; however, Moffett and Sarkar (16) provide a more complete discussion of the variety of multicriteria models.

All MCDM models can provide the ranking of alternatives, but none of them can be described to fit decision problems completely. For example, SAW uses a simple utility function model, which requires high-quality data (using real data or actual scores); however, decision makers may require only a ranking of the alternatives rather than their actual scores. Concordance methods do not require exact scores; therefore, they can effectively address criteria uncertainty. This method may provide only a partial ranking and preferred options rather than one best option. The selected MCDM methods should fit the complexity of problems, availability of data, and weighting technique. The authors provide an application of the SAW and ELECTRE III methods, which are assumed to fit with the available data and decision problems.

### Example Using "MCDM Models" in Step 5

The authors assume that following Steps 1 through 4, decision makers decide to use speed as a performance measure to assess the free-way performance at the implementation level. Decision makers are asked to define the criteria and weights to assess the data collection strategies. Assuming that a weighting technique is used to obtain the weight for each criterion and subcriterion shown in Figure 4, the three main criteria are composed of cost, accuracy, and reliability weighted by 65%, 20%, and 15%, respectively. The data processing time criterion is not considered because all equipment meets the criterion thresholds.

Five subcriteria are considered; however, the authors assume that decision makers are not concerned about the maintenance cost and human reliability criteria and assign them a zero for their weight.

On the basis of the screening of alternatives in Step 4, the six remaining alternatives are loop detectors, microwave sensor, video sensor, infrared sensor, acoustic sensor, and ITS probe vehicle. For the quantitative criteria, the ideal point concept proposed by Hwang and Yoon (17) is used for criteria normalization:

$$C_{ijk} = \frac{x_{ijk} - \min x_{ijk}}{\max x_{ijk} - \min x_{ijk}}$$
(1)

$$C_{ijk} = \frac{\max x_{ijk} - x_{ijk}}{\max x_{ijk} - \min x_{ijk}}$$
 (2)

where  $x_{ijk}$  is the score of alternative k with respect to criteria i on subcriteria j and  $C_{ijk}$  indicates the benefit of performance measure.

In the ideal point concept, the quantitative data are converted into a comparable unit between "0 and 1" where "0" is the lowest utility value. The use of Equations 1 and 2 depends on the variable  $x_{ijk}$ . If an increase of  $x_{ijk}$  leads to improvement, then Equation 1 is used; otherwise Equation 2 is applied. For example, if  $x_{ijk}$  means an increase in travel time, then a travel time increase will be unfavorable and Equation 2 is used. Equations 1 and 2 are applied in this example, and the scaling values are shown in Table 1. Even though the ideal point concept is simple and easy to apply, decision makers should be aware that small relative differences in the range of a criterion's values may lead to unrealistic criterion scaling. To counter that problem, the ELECTRE III method, which uses an indifference threshold, may be used. If the difference between criterion scores is less than the indifference threshold, the alternatives have the same performance level.

### SAW Method

In Figure 4, the decision hierarchies are composed of two levels: main criteria (level i) and subcriteria (level j). The form of an additive utility function in the upper level (i) is

$$V_{\text{tot}} = \sum_{i=1}^{n} w_{ik} U_{ijk} \tag{3}$$

where

 $V_{\text{tot}}$  = overall valuation for alternative k,

 $w_{ik}$  = weight assigned to criterion *i* for alternative *k*,

 $U_{ijk}$  = utility in the lower level j for alternative k to criterion i, and n = number of criteria.

The utility function in the lower level is calculated as follows:

$$U_{ijk} = \sum_{i=1}^{m} w'_{ijk} C_{ijk} \tag{4}$$

where

 $U_{ijk}$  = utility in the lower level j for alternative k to criterion i, m = number of subcriteria of criterion j,

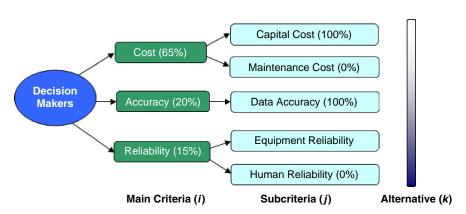


FIGURE 4 Weighting criteria and subcriteria.

ITS Probe Microwave Video Infrared Acoustic Subcriteria Sensor<sup>b</sup> Loop Detector Sensor Sensor Sensor<sup>d</sup> Vehicle 13,000 Capital cost (\$) 14,400 13,000 20,000 5,600 100,000 Data accuracy (%) 95 95 95 90 90 90 92.5 95 95 95 95 Equipments reliability (%) 96 Scaling subcriteria using Equations 1 and 2 Capital cost (\$) 0.91 0.92 0.92 0.85 1.00 0.00 Data accuracy (%) 1.00 1.00 1.00 0.000.000.00Equipments reliability (%) 0.71 0.71 0.71 0.71 0.001.00

TABLE 1 Scaling Alternative Criteria Scores

NOTE: All data are assumed at a freeway segment, approximately 2 mi. ITS = intelligent transportation system.

 $w'_{ijk}$  = weight assigned to subcriterion j to criterion i, and  $C_{ijk}$  = scaling value calculated from either Equation 1 or Equation 2.

The final decision is based on the result from the overall valuation in Equation 3.

#### ELECTRE III Method

The ELECTRE III method uses the concept of a concordance and discordance index to obtain alternative rankings. A concordance index C(a, b) for any pair of alternatives implies that alternative a is at least as good as alternative b. Then, the concordance index, C(a, b), is calculated as follows:

$$C(a,b) = \sum_{j=1}^{n} w_j c_j(a,b)$$
(5)

where  $w_j$  is the relative importance of the different criteria and  $c_j(a, b)$  is the local concordance index and can take values from 0 to 1.

$$c_{j}(a,b) = \begin{cases} 1 & \text{if } g_{j}(a) + q_{j} \ge g_{j}(b) \\ & \text{alternative } b \text{ is not preferred} \end{cases}$$

$$0 & \text{if } g_{j}(a) + p_{j} \le q_{j}(b) \\ & \text{alternative } b \text{ is preferred}$$

$$\frac{g_{j}(a) - g_{j}(b) + p_{j}}{p_{j} - q_{j}} & \text{otherwise}$$

$$(6)$$

where  $g_j(a)$  and  $g_j(b)$  are the performance scores of criterion j for alternative a and b and  $p_j$  and  $q_j$  are the preference and indifference thresholds for criterion j.

The discordance index is used to model the magnitude of the lack of compensation between the criteria by using a veto threshold  $v_j$  (constant threshold), which is set to check whether one alternative is very much preferred to another. If such a case occurs, the credibility index will be adjusted to decrease the global concordance index. A discordance index  $D_j(a, b)$  for each pair of alternatives implies no alternative a is better than alternative b. Then, the discordance index is calculated as follows (7):

$$D_{j}(a,b) = \begin{cases} 0 & \text{if } g_{j}(a) + p \ge g_{j}(b) \\ & \text{alternative } b \text{ is not preferred} \end{cases}$$

$$D_{j}(a,b) = \begin{cases} 1 & \text{if } g_{j}(a) + v_{j} \le g_{j}(b) \\ & \text{alternative } b \text{ is very much preferred} \end{cases}$$

$$\frac{g_{j}(b) - g_{j}(b) - p_{j}}{v_{j} - p_{j}} \quad \text{otherwise}$$

$$(7)$$

The discordance indices of different criteria are not aggregated using the criteria weights because each criterion is evaluated for discordance individually. The degree of outranking or credibility index is defined as follows:

$$S(a,b) = \begin{cases} C(a,b) & \text{if } D_j(a,b) \le C(a,b) & \forall j \\ C(a,b) \prod_{j \in J(a,b)} \frac{1 - D_j(a,b)}{1 - C_j(a,b)} & \text{otherwise} \end{cases}$$
(8)

where J(a, b) is a set of criteria for which  $D_i(a, b) > C(a, b)$ .

The credibility index is used to assess the trade-off between alternatives a and b. Alternative a will outrank alternative b when S(a, b) is greater than a minimum "threshold" value,  $\lambda$ , which is usually set at approximately 0.85, and S(a, b) minus S(a, b) is greater than a minimum threshold value, s, usually set at approximately 0.15. Then, a positive score +1 will be given to alternative a. In contrast, a negative score a1 will be given to alternative a2 being outranked. The final ranking will be established based on the total score through the process of descending and ascending distillation.

The value of the preference threshold (p) and indifference threshold (q) is set as the margins of uncertainty, error, or imprecision (14). The p and q threshold can be defined according to the decision makers' opinions. The p threshold is related to the positive attitude that a decision maker may have for a particular criterion's score. In addition, the q threshold is the point at which decision makers perceive a difference between alternatives (18). The veto threshold (v) can be set against the hypothesis that alternative a will usually be better than b. However, sometimes alternative a may be worse, or alternative b outperforms alternative a by at least the veto threshold. Thus, the veto threshold (v) must be greater than the p threshold.

<sup>&</sup>lt;sup>a</sup>Approximate cost for four lanes.

<sup>&</sup>lt;sup>b</sup>Covers up to eight lanes wide for one detector.

One camera can cover approximately four lanes.

<sup>&</sup>lt;sup>d</sup>One detector can cover approximately five lanes.

In this example, performance scores  $g_j(a)$  and  $g_j(b)$  use the scaled criterion scores in Table 1. The veto threshold is assumed to be equal to zero, and decision makers perceive the value of the p and q threshold as follows:

- Capital cost—this is an important criterion with maximum weight because it relates directly to the availability of types of equipment, techniques, and so forth. Thus, the q threshold is set to a "small" value (q=0.15), and the preference threshold is set twice as large (p=0.30). However, when using Equation 2 and the maximum and minimum value of the cost of capital equipment in Table 1, decision makers cannot distinguish the difference in the cost of equipment if it is less than  $[0.15 \times (100,00-5,600)] = \$14,160$ . In addition, decision makers can pay extra money, approximately  $[0.3 \times (100,00-5,600)] = \$28,320$ , for any equipment they prefer.
- Data accuracy and equipment reliability—this relates to the quality of information obtained. However, current technologies lead to only slight differences in equipment accuracy and reliability. Thus, the q threshold is set to a "large" value (q = 0.25), and the preference threshold is set twice as large (p = 0.50). However, decision makers cannot distinguish the difference between the performance of different types of equipment if the percent accuracy and reliability are less than 1.25% and 0.875%, respectively.

#### **Analysis Results**

The full ranking of alternatives using the SAW model based on the scaling data (Table 1) and weighting criteria (Figure 4) are analyzed. From this analysis, microwave and video sensors appear to be the best alternative, both having the highest score (0.91) using the SAW method. Loop detector, infrared sensor, acoustic sensor, and the intelligent transportation system ITS probe vehicle trail behind in that order. The ELECTRE III model is used to assess the same alternatives, and microwave and video sensors are again the best alternatives at the highest level; however, the loop detector is also one of the top alternatives (Figure 5).

The rank order for the two techniques is very similar; however, ELECTRE III does not provide a score for the alternatives as SAW does. This difference between the two techniques is a critical one

for decision makers and coordinators to consider. Recall specifically the earlier discussion concerning the decision statement in Step 1; the choices made at this stage will most likely determine the appropriate technique to apply in Step 5. The core concept behind the difference in the two techniques is ELECTRE III's assumption that small differences between alternatives are indistinguishable from one another because of inherent uncertainties in the decision-making process. As seen in Figure 5, the alternatives that share the same rank in ELECTRE III all have very similar scores in the SAW technique.

#### CONCLUSION

The framework provided in this paper improves on earlier research by facilitating the constrained selection of freeway operational performance measures. The process allows decision makers to generate candidate solutions and criteria before using qualitative and quantitative performance thresholds to reduce them to a set of feasible alternatives. The use of MCDM strategies provides an opportunity for decision makers to evaluate the trade-offs across the different criteria. Decision makers must select the appropriate weighting techniques or MCDM strategies based on the complexity of their problem or the required results, because each technique and MCDM strategy has strengths and limitations. The proposed criteria and application framework provides the guidelines for future applications by any agency. The successful implementation of the proposed methodology requires complete and engaged participation from decision makers.

Future research will use actual decision maker input to generate realistic criteria and their weights. In addition, decision maker input is required to establish the proper performance thresholds. Testing the process with decision maker participation will facilitate revision of the framework to fully meet decision makers' needs. After this framework is used for selecting operational freeway performance measures, it can be modified so that it is able to evaluate different operational strategies and recommend alternatives for selection. This future improvement is critical for the selection of operational alternatives in real time based on their associated performance measures.

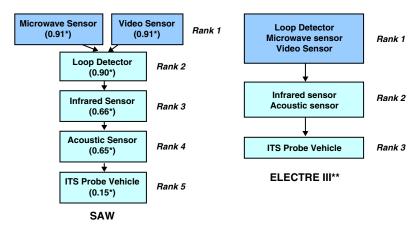


FIGURE 5 Ranking of alternatives by using SAW and ELECTRE III methods. (\*The total scores for each alternative based on Equations 3 and 4. \*\*Ranking results of descending distillation are the same as ascending distillation and final frequency.)

#### **ACKNOWLEDGMENTS**

The authors thank the Texas Department of Transportation for funding. The authors also thank the anonymous reviewers for their comments during the development of this work.

#### **REFERENCES**

- Freeway Management and Operations Handbook. Final Report. FHWA, U.S. Department of Transportation, Sept. 2003.
- Highway Capacity Manual. TRB, National Research Council, Washington, D.C., 2000.
- Shaw, T. NCHRP Synthesis of Highway Practice 311: Performance Measures of Operational Effectiveness for Highway Segments and Systems. Transportation Research Board of the National Academies, Washington, D.C., 2003. www.trb.org/publications/nchrp/nchrp\_syn\_ 311.pdf. Accessed April 4, 2008.
- Rogers, M. G. Engineering Project Appraisal: The Evaluation of Alternative Development Schemes. Blackwell Science Ltd., 2001.
- Brydia, R. E., W. H. Schneider, S. P. Mattingly, M. L. Sattler, and A. Upayokin. Operations-Oriented Performance Measures for Freeway Management Systems: Year 1 Report. Technical Report 0-5292-1. 2007. www.tti.tamu.edu/documents/0-5292-1.pdf. Accessed July 29, 2007.
- Office of Management and Budget. Performance Measurement Challenge and Strategies. June 2003. www.whitehouse.gov/omb/part/challenges\_ strategies.html. Accessed June 26, 2007.
- Jiji, D. Guide for Good Performance Measure (Beyond S.M.A.R.T.\*).
   Oct. 2005. www.i95coalition.org/PDF/Library/PerformanceMeasures/ Good%20Measures%20Guide-10-25-05%20from%20Dan%20Jiji.pdf. Accessed June 26, 2007.

- 8. Ababutain, A. Y. A Multi-Criteria Decision-Making Model for Selection of BOT Toll Road Proposals Within the Public Sector. PhD thesis. University of Pittsburgh, 2002.
- Dalkey, N. C., and O. Helmer. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, Vol. 9, 1963, pp. 458–467.
- National Transportation Operation Coalition (NTOC) Performance Measurement Initiative. Final Report. July 2005. http://www.ntoctalks. com/ntoc/ntoc\_final\_report.pdf. Accessed multiple dates, April through September 2006.
- Polatidis, H., D. A. Haralambopoulos, G. Munda, and R. Vreeker. Selecting an Appropriate Multi-Criteria Decision Analysis Technique for Renewable Energy Planning. *Energy Sources*, Part B, Vol. 1, 2006, pp.181–193.
- 12. Keeney, R., and H. Raiffa. *Decision with Multiple Objectives: Preferences and Value Trade-offs.* John Wiley & Sons, New York, 1976.
- Roy, B., and P. Vincke. Multicriteria Analysis: Survey and New Directions. *European Journal of Operational Research*, Vol. 8, 1981, pp. 207–218.
- Roy, B., M. Present, and D. Silhol. A Programming Method for Determining Which Paris Metro Stations Should Be Renovated. *European Journal of Operational Research*, Vol. 24, 1986, pp. 318–334.
- 15. Vincke, P. Multicriteria Decision Aid. Wiley, New York, 1992.
- Moffett, A., and S. Sarkar. Incorporating Multiple Criteria into the Design of Conservation Area Networks: A Mini-Review with Recommendations. *Diversity and Distributions*, Vol. 12, 2006, pp. 125–137.
- Hwang, C. L., and K. Yoon. Multiple Attributes Decision Making— Methods and Applications. Springer-Verlag, Berlin, 1981.
- Goletsis, Y., J. Psarras, and J. Samouilisdis. Project Ranking in the Armenian Energy Sector Using a Multicriteria Method for Groups. *Annals of Operations Research*, Vol. 120, 2003, pp. 135–157.

The Performance Measurement Committee sponsored publication of this paper.