Quantitative Methods for Design-Build Team Selection

Mounir El Asmar[1]; Wafik Lotfallah[2]; Gary Whited, M.ASCE[3]; and Awad S. Hanna, F.ASCE[4]

Abstract: The use of design/build (DB) contracting by transportation agencies has been steadily increasing as a project delivery system for large complex highway projects. However, moving to DB from traditional design-bid-build procurement can be a challenge. One significant barrier is gaining acceptance of a best-value selection process in which technical aspects of a proposal are considered separately and then combined with price to determine the winning proposal. These technical aspects mostly consist of qualitative criteria, thus making room for human errors or biases. Any perceived presence of bias or influence in the selection process can lead to public mistrust and protests by bidders. It is important that a rigorous quantitative mathematical analysis of the evaluation process be conducted to determine whether bias exists and to eliminate it. The paper discusses two potential sources of bias—evaluators and weighting model—in the DB selection process and presents mathematical models to detect and remove biases should they exist. A score normalization model deals with biases from the evaluators; then a graphical weight-space volume model and a Monte Carlo statistical sampling model are developed to remove biases from the weighting model. The models are then tested and demonstrated using results from the DB bridge replacement project for the collapsed Mississippi River bridge of Interstate 35W in Minneapolis.

DOI: 10.1061/(ASCE)CO.1943-7862.0000194

CE Database subject headings: Monte Carlo method; Design/build; Selection; Quantitative analysis; Bids; Construction industry.

Author keywords: Monte Carlo method; Design/build; Selection; Quantitative analysis.

Introduction

The use of design/build (DB) contracting by transportation agencies has been steadily increasing over the last decade and is becoming more popular as a project delivery system for large complex highway projects (Gransberg and Molenaar 2003; Contract Administration 2002). Transportation agencies are finding that the use of DB contracting can reduce the overall project delivery time with little impact on quality. However, moving to DB from the traditional design-bid-build (DBB) procurement can be a challenge. There are many actual and perceived barriers that must be overcome by both the owner agencies and the contracting community. One significant barrier can be getting a buy-in on the use of a best-value selection process with DB rather than the customary low-bid selection with DBB.

Public sector transportation projects tend to use selection methods in which technical aspects of the proposal are considered separately from the price component. The technical scores and price are then combined in some simple manner to determine the best value and the winning proposal (Herbsman and Ellis 1992). Technical scores are determined by a technical evaluation team on the basis of prestated criteria. The absence of a quantitative methodology for the technical evaluation, and the potential for bias or political influence to enter the process, have caused some of the greatest concern. Any perceived presence of bias or influence in the selection process can lead to public mistrust and protests by bidders (Shane et al. 2006). The case study presented in this paper described a project where the DB team selection and subsequent protest resulted in a great deal of public commentary and concern over the process. The following quotation from Engineering News Record describes the perceived presence of bias: "Two teams bidding on a replacement to the fallen I-35W bridge in Minneapolis have filed an administrative protest questioning the legality of Minnesota's design-build procedures for highway construction. The move also calls into question the subjectivity of best-value procurement models" (ENR, September 24, 2007).

Bias in the selection process can come from individual evaluators or from weighting criteria, and it can be difficult to detect by simply looking at the results. It is important that a more rigorous mathematical analysis of the evaluation process be conducted to determine whether bias exists and to eliminate it in order to have a fair and open selection process. Transportation agencies, the construction industry, and the public would have greater confidence in the DB best-value approach if the selection

[1]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin—Madison, 2256 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706 (corresponding author). E-mail: elasmar@wisc.edu

[2]Associate Professor, Dept. of Mathematics and Actuarial Science, American Univ. in Cairo, Egypt; and, Visiting Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin—Madison, 2208 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706. E-mail: lotfalla@math.wisc.edu

[3]Program Manager, Construction and Materials Support Center, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin—Madison, 2314 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706. E-mail: whited@engr.wisc.edu

[4]Professor and Chair, Construction Engineering and Management Program, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin—Madison, 2320 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706. E-mail: hanna@engr.wisc.edu

Note. This manuscript was submitted on December 12, 2008; approved on January 25, 2010; published online on January 28, 2010. Discussion period open until January 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 136, No. 8, August 1, 2010. ©ASCE, ISSN 0733-9364/2010/8-904–912/$25.00.

process were transparent and could be shown to be objective and unbiased.

Objective

The problem in DB team selection usually comes from combining different criteria, especially when some of them are qualitative. Misunderstanding the selection criteria can lead to protests, negative publicity and possibly litigation, especially in the public sector. The purpose of this paper is to enhance the existing selection approaches by applying mathematical methods to remove biases from the evaluation, providing a robust scientific selection methodology. The models are illustrated through a case study of the evaluation process of the bridge replacement project for the Mississippi River crossing of Interstate 35W in Minneapolis.

Scope

This paper assumes that the selection criteria are independent and not correlated. It is also assumed that the evaluators record their evaluations individually as opposed to a group consensus. The study is limited to DB selection in the United States. The developed methodology is tested on models that rely on simple averaging of different evaluation scores, which represents a large fraction of the industry. In fact, Gransberg et al. (2007) studied the selection strategies used in the highway construction industry; the results of the study show that only 37% of the industry awards projects to the lowest bidder that meets specific technical criteria. Most of the remaining 63% take into account cost, time, and quality and combine them using various equations specific to the individual agencies. This study only considers selection methodologies where different criteria are evaluated by several evaluators, and the total score is a weighted average of these individual criteria's scores.

Literature Review

The multicriteria selection problem has been studied for several industries, such as research and development project selection (Martino 1995) or engineering projects selection (Dean and Nishry 1965). Several researchers have studied multicriteria evaluation of construction contractors: Russell (1992) developed a "hybrid" decision model for contractor prequalification; Hatush and Skitmore (1998) developed an additive model using multicriteria utility theory; and Fong and Choi (2000) developed a final contractor selection model that uses the analytical hierarchy process. Singh and Tiong (2005) developed a "fuzzy decision framework for contractor selection."

Most of these studies were tailored to the traditional DBB project delivery system, taking into account that the design is ready and the contractors are bidding on a finished design. However, in DB, the bidding takes place at a very early stage and the DB teams are being evaluated on their capability of performing on the project. Nevertheless, selection criteria considered for DB projects can be very similar to criteria used for contractor selection for DBB projects, depending on the nature of the project. Molenaar and Songer (1998) showed that the best way to select a DB team for public projects is to take both price and quality into account during the selection process. The different selection criteria depend on the owner agency's requirements for the specific project, and studies have identified the most common criteria used and related project success to these criteria (Hatush and Skitmore 1997). A number of these criteria were listed by Gransberg and Ellicott (1997).

Gransberg and Senadheera (1999) listed the three common DB contract award methods used for transportation projects as low-bid DB, adjusted score DB, and the best-value DB. Beard et al. (2001) categorized the procurement methods for DB projects; El Wardani et al. (2006) compared these procurement methods to help decide on the most appropriate one to be used during selection of the DB team. The Utah DOT documented use of the best-value selection process for DB projects (Postma et al. 1999). Palaneeswaran et al. (2003) proposed a model for best-value contractor selection. Phipps (2000) documented the DB selection process for a Maine bridge, which can serve as an example of the typical selection methods used today.

Shane et al. (2006) stressed the legal challenges with the use of a procurement system that takes into account several technical factors, not only the lowest bid. This raises the issue of a need for a scientific robust model to assess the selection system used. The aim of the proposed methodology is to fill this gap to ensure that selection is unbiased and representative of a fair process.

Selection Methodology: General Model

This paper introduces a general model that can be applied to any DB team selection process, which selects the best possible team based on different criteria, with partial scores given by different evaluators. However, for illustration purposes, the model used for this paper is the DB best-value process presented next. The team selection approach combines three components (cost, time, and quality) into comparable scores for the competing teams. Time can either be converted to dollar amounts, and added to the cost of the project, or be evaluated and graded like the remaining criteria and added to the technical score.

The model assigns to each team j a corresponding contract bid price (cost) P_j, and a technical score (quality) S_j (one of which should account for time, depending on the owner agency). The selection equation used to compute the final score (or adjusted bid) F_j is $F_j = P_j / S_j$ (which may be thought of as the price per quality unit), and the contract will be awarded to the lowest adjusted bid.

The technical score S_j is based on several evaluation criteria with various weights. To describe the algorithm for evaluating S_j, we use the following notation: let j denote team number j, $1 \leq j \leq J$; k=evaluator number, $1 \leq k \leq K$; w_i=weight of criterion i, $1 \leq i \leq I$; and s_{ijk}=score of criterion i for team j given by evaluator k.

The current selection methodology used by many state highway agencies (SHAs) (Gransberg et al. 2007) consists of simply averaging the scores without considering any possible biases that might occur in the evaluation process. In this selection methodology, the technical score S_j is simply calculated as the weighted average of the different criteria, after taking the average of the different evaluators' scores. Thus: $S_j = \sum_{i=1}^{I} w_i [\sum_{k=1}^{K} (s_{ijk}) / K]$.

In order to increase the chances of the project success, we need to ensure that possible bias sources have been investigated and that any forms of unfairness have been removed from the evaluation. There are two types of potential biases, since there are two inputs to the model: scores and weights. The first source of bias is the evaluators themselves, who can, for example, be inclined toward one of the teams. The second source of bias comes from the model, as a high weight on a certain criterion can skew the results in favor of a specific team.

J. Constr. Eng. Manage., 2010, 136(8): 904-912

Table 1. Normalization Example

Weight (%)	Team 1	Evaluator 1	Evaluator 2	Evaluator 3	Original scores	Revised scores
30	Criterion 1	32	84	91	65.73	72.17
70	Criterion 2	28	80	85		
	Team 2					
30	Criterion 1	80	60	65	67.40	60.97
70	Criterion 2	83	58	60		

Evaluators' Biases: Normalizing the Scores

Evaluator's Power

When evaluating teams on a certain criterion, two evaluators may have the same opinion concerning the teams' respective qualifications. However, one might give team scores that differ by a small range and the other give scores that differ by a much larger range. For example, if one evaluator's range is one-half of another evaluator's, the latter will obviously have a greater influence on the overall score when scores for a specific team are averaged. In fact, one can show that the second evaluator will have twice the impact of the first evaluator on the final score. This is what the writers call the "evaluator's power."

If an evaluator understands this power, he/she can significantly influence the decision. Therefore, in order for the evaluators to be given equal powers, their scores should be normalized before any analysis is performed. This will transform their evaluations into comparable scores that can later be used to compute the average scores on which the selection decision is based. Different normalization methods were used in the literature, but the normalization method described here is tailored to suit this paper's particular application: DB team selection.

It should be noted that these models are applicable to cases where scores are individually and independently achieved. For instance, if scores are not individually achieved, and instead result from a partial or thorough discussion of the proposals among the evaluators, bias could actually be a group bias that would not be apparent in this diagnostic.

Normalization Technique

To facilitate the reading of the formulas, we use the following notations: the operation $\mathrm{Av}_k(X_{ijk})=(1/K)\Sigma_{k=1}^{K}X_{ijk}$ denotes the average of the n-dimensional array X_{ijk} over the index k, fixing the other indices. Also, $\min_j(X_{ijk})=\min\{X_{ijk}:j=1,\ldots,J\}$ denotes the minimum of the array X_{ijk} over the index j, fixing the other indices; similarly, $\max_j(X_{ijk})$ is used for the maximum.

The normalization procedure maps each score s_{ijk} to a normalized version z_{ijk} as follows: fixing a criterion i and an evaluator k, the teams are varied, and their mean score and standard deviation are calculated. These are

$$\mu_{ik}=Av_j(s_{ijk}) \text{ and } \sigma_{ik}=\sqrt{Av_j((s_{ijk}-\mu_{ik})^2)}, \text{ respectively}$$

The z-scores are then calculated to be $z_{ijk}=(s_{ijk}-\mu_{ik})/\sigma_{ik}$. These z-scores hide any variation of the ranges of different evaluators, giving them equal powers. At this point we can calculate the average z-score for each criterion i and each team j by the equation $a_{ij}=Av_k(z_{ijk})$. We note that the averages a_{ij} of the z-scores are centered around zero; some values are positive and some are negative. To enter the final equation, these scores need to be mapped back to numbers r_{ij} (called the revised scores) which fit the original scale of evaluations, a 100-point scale. This procedure, called denormalization, should take place for two reasons.

First, scores a_{ij} are useless in the final equation because one cannot adjust price through dividing by a negative number (or zero). Second, the average of the ranges given by different evaluators over the same criterion should be respected and considered as the actual range of the scores of the different teams on this particular criterion. The reason for this is that evaluators may agree that all teams have the same quality in some criterion, while they differ on another.

Thus, the working range needs to be set for every criterion. Data from the specific selection case is used to find this range. The study uses the average of the lowest scores and the average of the highest scores for every criterion. This is because the evaluators "agreed" on these ranges when scoring the teams, and the scores are considered to reflect the "wisdom" of the evaluators. Thus, it is natural to have the revised scores r_{ij} in this range.

Going back to the model's original scores and z-scores, the writers fix the evaluators and criteria, only varying the teams and take their minimum score. The average of these minimum scores over the evaluators is calculated for every single criterion and is $L_i^s=\mathrm{Av}_k[\min_j(s_{ijk})]$. The same procedure is performed for the maximum values to get their average $H_i^s=\mathrm{Av}_k[\max_j(s_{ijk})]$. L_i^z and H_i^z are calculated for the z-scores in the same manner, replacing s_{ijk} by z_{ijk}. With this done, the mapping equation gives us the following new revised scores:

$$r_{ij}=L_i^s+\frac{H_i^s-L_i^s}{H_i^z-L_i^z}(a_{ij}-L_i^z)$$

Thus, revised scores r_{ij} are simply the output of a linear function of a_{ij} mapping L_i^z to L_i^s, and H_i^z to H_i^s. These revised scores can be used in the weighted average equation $S_j=\Sigma_{i=1}^{I}w_ir_{ij}$ that will be set as the denominator to get the adjusted bid.

The following simple numerical example will demonstrate how normalization of the scores removes bias from the evaluation. Consider three evaluators selecting one of two competing teams, according to two criteria, weighted 30% and 70%, respectively. Table 1 shows the calculated original and revised scores.

If the straightforward weighted average method is used (original scores), Team 2 will score higher. However, if one looks at the scores, one notices that Evaluator 1 could be biased for Team 2, since he/she disagrees with the other evaluators, who preferred Team 1. Moreover, Evaluator 1 used a much wider range of scores than the other two, thus tilting the results in favor of Team 2. By using the normalization technique presented here, it would not matter who is biased anymore, since all evaluators will be given equal powers. The writers understand that those individual scores could be very well justified, depending upon the perspective of the individual evaluator, and this is why we should not just drop some evaluator, if his/her scores change the outcome of the process. Instead, variations in the scoring of a criterion should be adjusted by the average of the variations of all evaluators on this particular aspect to give all evaluators equal powers while still factoring their views in the equation. The majority of the evaluators are assumed to be fair; their scores will offset the original biased scores.

Note that the case presented here is simple (only two criteria and three evaluators) and it was easy to see the potential bias by just looking at the scores. However, when there are many evaluators and several criteria, bias cannot be spotted easily. On the other hand, the use of the revised scores in the discussed framework removes the bias from Evaluator 1 and Team 1 gets the highest score.

Testing for Irregularities

For more reliable results, in the presence of four or more evaluators, one can test for possible unfairness or bias due to a particular evaluator. Thus, when calculating the z-scores z_{ijk}, the model should be run several times after the removal of each evaluator in turn, while monitoring results to identify changes due to possible biased evaluators.

Another source of bias is the presence of extreme scores given by (possibly) more than one evaluator. This bias can be eliminated for each criterion, by removing both the minimum and maximum z-scores before calculating the average score α_{ij}.

Testing and Validating the Weight Model

Weight-Space Model

The weighting of the evaluation criteria could be set in such a way to favor a particular contractor, whose point of strength is already known. This is true since the contractors cannot do much to improve their possible scores with respect to some criteria. An example would be when a design/builder is known to design very artistic and aesthetically appealing bridges; this team can be favored by setting a very high weight on the aesthetics criterion, with respect to how important that is for the project.

A mathematical model was also developed to show how much the weights can vary without affecting the selection decision. The range, in which the weights can vary without affecting the outcome, shows the amount of robustness in the weighting system. A particular list of weights $w_1, w_2, \ldots, w_I$ given to the different criteria can be visualized as a point in the Ith dimensional space. However, since weights must sum up to 1, there are $(I-1)$ free weights, a point in the $(I-1)$th dimensional space. It turns out that this space is divided into regions separated by hyperplanes, with each region representing a winning region for a particular team. One can draw a hypersphere centered at the working point (=assigned list of weights) inside the region of the winning team. This hypersphere can grow with no change in the winning team, up to a point where it reaches the closest hyperplane, outside which the decision will differ.

The first two conditions that have to be met by the weights are the following:

- Weights are nonnegative, $w_i \geqslant 0$; this puts us in the first octant when translated graphically in the three-dimensional (3D) space; and
- The sum of all weights is 100%, i.e., $\Sigma_{i=1}^{I} w_i = 1$. Thus, considering only the $(I-1)$ free variables $w_1, w_2, \ldots, w_{I-1}$, we get the inequality $w_I = 1 - \Sigma_{i=1}^{I-1} w_i \geqslant 0$, i.e., $\Sigma_{i=1}^{I-1} w_i \leqslant 1$.

Thus, in the 3D illustration, our working region is a pyramid in the first octant. To find the separating hyperplanes we consider the scores

$$S_j = \sum_{i=1}^{I-1} r_{ij} w_i + r_{Ij}\left(1 - \sum_{i=1}^{I-1} w_i\right)$$

where r_{ij}=revised scores found in the previous section. Instead of dividing S_j by P_j in order to get F_j, the total score for T_j, we consider its reciprocal

$$1/F_j = c_j + \sum_{i=1}^{I-1} b_{ij} w_i$$

where $c_j = s_{Ij}/p_j$ and $b_{ij} = (s_{ij} - s_{Ij})/p_j$.

Knowing that teams are trying to achieve the lowest possible total score, let us compare two teams j and j' where T_j is the winning team

$$F_j < F_{j'} \Leftrightarrow c_j + \sum_{i=1}^{I-1} b_{ij} w_i > c_{j'} + \sum_{i=1}^{I-1} b_{ij'} w_i$$

$$\Leftrightarrow c_j - c_{j'} + \sum_{i=1}^{I-1} (b_{ij} - b_{ij'}) w_i > 0$$

The number of inequalities/conditions that separate the teams is dependent on the number of teams J being evaluated, and is the combination

$$\binom{J}{2} = J(J-1)/2$$

where J=number of teams. Distances from the working point to every one of the hyperplanes/boundaries can be measured in weight percentages.

Letting $n = I-1$, the working point is designated by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and a condition represented by a hyperplane with the equation $b_1 w_1 + \cdots + b_n w_n + c = 0$, the distance between the point and the hyperplane can be found to be

$$d = |b_1 x_1 + \cdots + b_n x_n + c| / \sqrt{b_1^2 + \cdots + b_n^2}$$

The smallest distance d to the closest hyperplane will show how much these weights can vary without affecting the initial results/decision. A larger distance shows that the model is more robust. Distances calculated will be interpreted as the length of the vector of possible differences, i.e.

$$d = |\Delta \mathbf{w}| = \sqrt{\sum_{i=1}^{I-1} (\Delta w_i)^2}$$

As an example, let us consider the special case where all the weights vary within the same percentage, i.e., $|\Delta w_i| = c$ (a constant); then $d = \sqrt{(I-1)(\Delta w_i)^2} = c\sqrt{I-1}$, which is equivalent to $|\Delta w_i| = d/\sqrt{(I-1)}$.

The projection of the working point (of weights) on each of the hyperplanes can also be calculated. This will give the combinations of weights closest to the working point that will result in a tie between the two teams separated by the hyperplanes. Let $\mathbf{x}$ be (the position vector of) the working point, and the plane equation $b_1 w_1 + \cdots + b_n w_n + c = 0$ represented by $\mathbf{b} \cdot \mathbf{w} + c = 0$ (where $\cdot$ is the dot product). Then the projection $\mathbf{y}$ will be

J. Constr. Eng. Manage., 2010, 136(8): 904-912

Table 2. Weight-Space Example

	Weight (%)	Team 1	Team 2	Team 3
Criterion 1	20	90	50	50
Criterion 2	30	50	60	80
Criterion 3	20	60	80	70
Criterion 4	30	70	90	60
Total	100	66	71	66

$$\mathbf{y}=\mathbf{x}-(\mathbf{b}\cdot\mathbf{x}+c)\frac{\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}}$$

It can also be noted that the distance $d=|\mathbf{y}-\mathbf{x}|$.

Let us take the scores in Table 2 as an example. The evaluators are not present here since the model is being analyzed after the denormalization step. Three teams are being evaluated on four criteria. The number of criteria was restricted to four for illustration purposes.

In a 3D graph, the working point of this example is $\mathbf{x}=(0.2,0.3,0.2)$, with the fourth weight being (1−the sum of the other three). A total of four plane equations can be derived from this model; three divide the winning teams and the last represents the condition under which the sum of weights should be less than or equal to 1. Substituting w_4 by $w_4=1-w_1-w_2-w_3$, the equations are

$$w_3=-6w_1+2 \tag{1}$$

$$w_3=-0.6w_1-0.4w_2+0.6 \tag{2}$$

$$w_3=0.75w_1-0.5w_2+0.25 \tag{3}$$

$$w_1+w_2+w_3=1 \tag{4}$$

Team 2 wins in this example. The distance between $\mathbf{x}$ and Plane (1) was found to be 9.86% whereas the distance between $\mathbf{x}$ and Plane (2) was 20.32%. This means that the weights can vary by $\Delta w_i=9.86\%/\sqrt{(4-1)}=5.69\%$ each without affecting the decision. In the graphical representation of the model below (Fig. 1), the planes (1), (2), and (4) and the working point $\mathbf{x}$ are shown; the sphere has the radius $d=9.86\%$.

Fig. 1 shows the white-colored plane that represents the condition under which the sum of all weights is equal to 100%. The shaded planes divide the winning regions for the three competing teams. The working point is represented by the black dot in Team 2's winning region, which means that Team 2 is the winner under the current weight matrix. The sphere centered around this working point reflects the fact that Team 2 will still be the winner even when any criterion weight is changed by up to about 6% in this specific example.

It should be noted that the computational complexity of the weight-space model presented in this paper is relatively low. The reason behind this is that the calculation of the distances between the working point and the separating hyperplanes uses simple and straightforward formulas, independent of the space dimension.

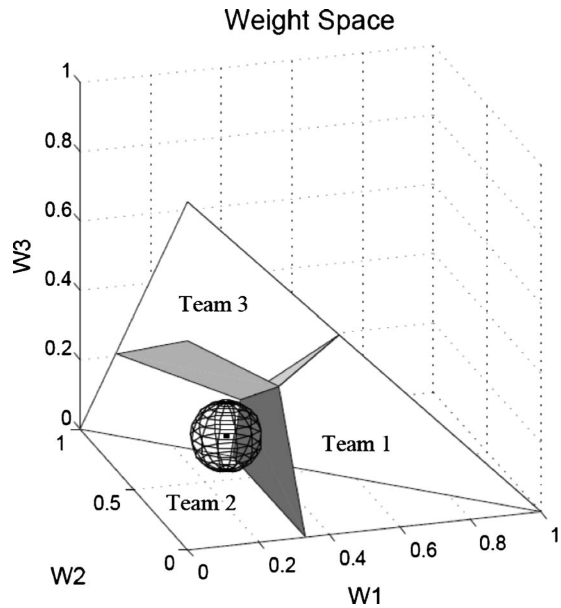

Fig. 1. Weight-space example

Monte Carlo Simulations: Finding the Volumes of the Winning Regions

If the working point is close to a hyperplane, and thus a hypersphere with a reasonable radius will have a portion contained in a different region, one may try to evaluate the proportion of the volume of the hypersphere (or of a hyperparallelepiped) that is contained in the region of the winning team.

Since evaluating volumes in higher dimensional spaces involves complicated multiple integrations, a much easier method is to use a Monte Carlo simulation, which repeatedly randomly assigns weights according to a specified distribution for every single criterion, and recalculates the total score for every team, then compares these scores to identify the winner.

Starting with N weights, to guarantee a uniform distribution on the space of feasible weights (those that add up to 1), we choose random values (based on a uniform distribution with symmetric intervals around the initial values) for $N-1$ weights, then calculate the nth weight from $1-\Sigma_{i=1}^{N-1}w_i$. The idea of randomly choosing n weights (say in the interval $[0,1]$) and normalize their values to add up to 1, looks natural, but will produce a nonuniform distribution that favors the values that are close to $1/n$. This can be simply seen using a geometrical argument by mapping the unit square onto the line segment $y=1-x, 0\leqslant x\leqslant 1$, where each point (x_0,y_0) in the unit square maps to the intersection of the line $y=(y_0/x_0)x$ with the line $y=1-x$.

The number of times needed for the results to converge is calculated as follows. Let the probability for team j to win be $\Pr(j)$. Using a uniform distribution, this probability represents the proportion of volume of the hyperparallelepiped inside the region of team j. Also, let N be the total number of iterations and $X_i, i=1,\ldots,N$ be the random variables defined by

$$X_i=\begin{cases}1, & \text{if the original winning team still wins}\\ 0, & \text{otherwise}\end{cases}$$

Then if N is large enough, the central limit theorem will result in $\Sigma_{i=1}^{N}X_i$ being approximately normally distributed with mean $\mu=N\cdot\Pr(j)$ and standard deviation $\sigma=\sqrt{N\cdot Pr(j)\cdot(\Pr(j))}=\sqrt{Npq}$ where $p=Pr(j)$ and $q=1-\Pr(j)$. Then $Av_i(X_i)$ would have a standard deviation $\sigma=\sqrt{Npq}/N=\sqrt{pq/N}$. The main objective is to decrease σ as much as possible. Since $q+p=1$, the worst case scenario would be when $q=p=0.5$ where σ would be the largest.

J. Constr. Eng. Manage., 2010, 136(8): 904-912

Let us consider that for the purpose of this study the required standard deviation is $\sigma = 0.5\%$. Thus, $N = (0.5/\sigma)^2 = (0.5/0.5\%)^2 = 10{,}000$ iterations. It must be remembered that this is only an illustration for the worst case scenario. However, more iterations can be performed, if needed, in order to get better accuracy.

For this study, uniform distributions were used (for calculation of volume proportion). To set the endpoints of the distribution, the minimum value is set to zero because the weights of the criteria cannot be negative. Maximum values are hence set to double the original value. The reason is that the owner agency's requirements are to be taken into consideration, i.e., the range should be centered around the original value to represent the owner's needs for the specific project. Also, within this restriction the maximum possible range is sought.

For example, when a criterion weighs 10% of the total score in the original model, the simulation will pick random values between 0 and 20% to see how this affects the results. This is done for every single criterion in turn, and the total score is calculated using these new weights. Because of the wide spectrum of weights that one can have, it is imperative that numbers be realistic. Hence, conditions for the run to be registered are that all the free weights are positive and their sum is less than 100%.

Case Study

Introduction and Background

On August 1, 2007, the I-35W bridge over the Mississippi River in Minneapolis collapsed, resulting in the tragic loss of 13 lives. The Minnesota DOT (MN/DOT) chose to use a DB procurement process to design and build a replacement bridge as quickly as possible.

MN/DOT used a two-phase process to select the DB contractor to deliver the project. The first phase involved issuing a request for qualifications, and resulted in five prequalified teams to respond to the request for proposals (RFP). Phase two involved issuing the RFP. Four teams ultimately submitted proposals on the project, and a contract was awarded based upon a best-value determination.

The best value was based upon cost plus time (adjusted) divided by the technical score. The technical score was determined by evaluating and scoring each proposal on predetermined criteria. Technical evaluations and scoring were done by a six-member Technical Review Committee that included members from MN/DOT, city of Minneapolis, and the Association of General Contractors.

Ultimately the best value was determined to be the proposal with the highest cost and the longest construction duration. This resulted in a great deal of public commentary and a formal protest by two of the submitters. The protest asserted that aspects of the scoring were arbitrary and capricious as well as citing other issues. The protest was ultimately denied after a thorough review of all assertions.

The I-35W bridge replacement project and DB process have undergone intense scrutiny, and MN/DOT is to be commended on the level of transparency provided. Scoring for each proposal has been made public and allows us to use the project as an example of how the evaluation models proposed in this research can assist transportation agencies in evaluating DB proposals and ensure that public trust is maintained through the selection process.

Selection Methodology Used for the I-35W Bridge Replacement Project

The team selection approach details can be found on the MN/DOT website. The selection followed the DB best value "$A+B$" process, which combines the three components, cost, time, and "quality," into comparable scores for the competing teams. The model's variables are as follows:

- "A" = contract bid price;
- "B" = number of days to complete the project, which will then be multiplied by \$200,000 per day (based on 50% of road user costs); and
- TPA = technical proposal average score.

and the selection equation is

$$\text{Adjusted bid} = \frac{A + (B \times \$200{,}000)}{\text{TPA}}$$

The contract was awarded to the lowest adjusted bid. The technical score was based on the following evaluation criteria:

1. Quality (50%)
   a. Experience and authority of key individuals (20%);
   b. Extent of quality control/quality assurance (10%);
   c. Safety (10%); and
   d. Measures to evaluate performance in construction (10%).
2. Aesthetics/visual quality (20%)
   a. Enhancement to the RFP (10%); and
   b. Approach to involve stakeholders (10%).
3. Enhancements (15%)
   a. Geometric enhancements (10%); and
   b. Structural enhancements (5%).
4. Public outreach/involvement (15%)

The selection criteria and evaluation model were known to bidders prior to proposal submittal (Minnesota DOT 2007). The writers understand that the quantification of time into dollar amounts can be an additional source of bias to the model, but the scope of the research is limited to removing bias from evaluators and weights. In order to avoid having to quantify time, owner agencies have other options available, such as adding time as a regular criterion and scoring it on a 100% scale just like the other criteria.

Revised Results

Evaluators' Biases—Normalizing the Scores

Scores for the individual criteria were all converted to a 100-point scale. A sample for one team's original scores (on a 100-point scale) and z-scores are shown in Tables 3 and 4. The remaining original scores tables can be found on the MN/DOT website, and could not be included in the paper for space limitations. The original scores and the revised scores are shown in Tables 5 and 6 for all four teams.

In this case study, the revised scores are very similar to the original scores, a result that confirms the absence of evaluators' biases in this project. If any evaluators participating to the selection were biased, the revised scores would have been different, reflecting this partiality. The second possible source of bias—the weights—is investigated next.

Weight-Space Model

In this case study, four teams are competing, so

J. Constr. Eng. Manage., 2010, 136(8): 904-912

Table 3. Sample Score for One of the Teams

Technical proposal	Weight (%)	Maximum potential points	Technical proposal score for Team 1					
			Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4	Evaluator 5	Evaluator 6
Quality (50%)								
Experience	20	100	60.00	40.00	70.00	50.00	63.00	60.00
Quality assurance/quality control (QC/QA)	10	100	55.00	45.00	72.00	60.00	60.00	50.00
Safety	10	100	80.00	79.00	91.00	88.00	96.00	80.00
Performance	10	100	50.00	55.00	71.00	65.00	51.00	55.00
Aesthetics (20%)								
Enhance RFP	10	100	75.00	68.00	85.00	75.00	52.00	55.00
Involve stakeholders	10	100	50.00	20.00	74.00	50.00	50.00	50.00
Enhancements (15%)								
Geometric	10	100	0.00	0.00	25.00	10.00	0.00	0.00
Structural	5	100	30.00	20.00	40.00	51.00	25.00	0.00
Public relations (15%)	15	100	73.00	78.00	84.00	62.00	55.00	80.00
Total	100	100	55.45	47.40	70.40	56.65	53.00	53.00
Score			55.98					
Price			$178,489,561.00					
Time (days)			392					
Time (days × $200,000)			$78,400,000.00					
Adjusted score			4,588,679.27					

$$\binom{4}{2} = 6$$

inequalities will be added to the conditions stated earlier. This gives a total of seven planes in the first octant. The model will be eight dimensional since the teams are being evaluated on nine criteria. The distances between the working point and the planes can be calculated. For the special case of all the weights varying by equal percentages, $\Delta x_i = d/2\sqrt{2} \approx d/2.83$ means that whatever d will be, all the weights can vary, each of them by $d/2.83$, without affecting the original decision.

An eight-dimensional model cannot be graphed. The writers had to choose a maximum of four criteria to graph, while fixing the other criteria. The four quality criteria were chosen. The reason for choosing the first criterion is that the range of scores between the winning team and the contestant is relatively high and its weight is the highest among all the others. So diminishing the weight of this criterion might give a chance to the contestant to win, if this difference offsets the difference in cost between them. The other three criteria were chosen because the teams' scores were relatively close to each other, which also give some chances to the other contestant when the cost is considered. If the other criteria were chosen, the contestants would not have high chances of winning. This model is hence trying to show the contestants' best likelihood of winning. In this case, fixing five criteria and varying the four quality criteria gives the graph that is shown in Fig. 2 below.

By looking at Fig. 2, the results show that only the plane that divides Teams 2 and 3 is in the study's working range. Team 3 is still the winner in most cases whereas Team 2 occupies a small volume relative to the winner. It should be noted that this combination of criteria represents Team 2's highest chances.

Table 4. Sample *Z*-Scores for One of the Teams

Technical proposal	*Z*-scores for Team 1					
	Evaluator 1	Evaluator 2	Evaluator 3	Evaluator 4	Evaluator 5	Evaluator 6
Quality (50%)						
Experience	−0.9298	−1.0370	−0.7040	−1.3081	−0.7477	−0.4834
QC/QA	−1.3778	−1.3757	−1.2362	−1.0824	−0.8956	−1.3472
Safety	0.9623	0.1350	0.9988	1.0835	1.6314	−0.8552
Performance	−1.5213	−1.2097	−0.7579	−0.9623	−1.2563	−1.6775
Aesthetics (20%)						
Enhance RFP	0.2027	0.1222	−0.1516	−0.2179	−1.0804	−0.8021
Involve stakeholders	−0.7385	−0.6765	−0.5347	−1.0675	−0.9467	−1.3484
Enhancements (15%)						
Geometric	−1.5669	−1.2268	−1.7126	−1.4614	−1.5703	−0.9526
Structural	−1.2347	−1.4597	−1.4995	−1.0553	−1.3374	−1.5966
Public relations (15%)	0.3922	0.2936	−1.0690	−0.6091	−1.7009	−0.3333

J. Constr. Eng. Manage., 2010, 136(8): 904-912

Table 5. Original Scores for All Four Teams

Technical proposal	Weight (%)	Original scores Team 1	Team 2	Team 3	Team 4
Quality (50%)					
Experience	20	57.17	60.33	94.17	71.67
QC/QA	10	57.00	71.17	88.00	70.50
Safety	10	85.67	76.33	86.00	74.33
Performance	10	57.83	77.83	79.00	66.33
Aesthetics (20%)					
Enhance RFP	10	68.33	63.67	97.83	63.83
Involve stakeholders	10	49.00	61.33	97.17	56.33
Enhancements (15%)					
Geometric	10	5.83	43.00	92.50	67.00
Structural	5	27.67	63.17	93.00	59.33
Public relations (15%)	15	72.00	75.67	92.33	71.67
Total	100	55.98	65.91	91.47	67.88
Price (in millions)		\$178.489	\$176.938	\$233.763	\$219.000
Time (days)		392	367	437	437
Time (days × \$200,000) (in millions)		\$78.4	\$73.4	\$87.4	\$87.4
Adjusted score (in millions)		4.589	3.798	3.511	4.514

Monte Carlo Simulations

In this case study, 10,000 iterations were performed to figure out approximate values for p and q, 7,796 of which were validated. The results were $p=0.986$ and $q=0.014$, giving a $\sigma=\sqrt{pq/N} = \sqrt{(0.986\times 0.014)/7{,}796}=0.133\%$, which is reasonable considering that the total scores differ by much more than this value.

The results were in accordance with the original scores. In fact, Team 3 won 94.3% of the time, and Team 2 won the remaining 5.7% of the time. Therefore, after the evaluators' scores were revised and probability distributions were added to the weights, the selection model used and the decision made were validated.

This scientific methodology resulted in the same outcome as the original case study. However, this might not always be the case. When this happens, the decision should be based on the developed model results, and not on the original weighted average technique, since the latter is only a special case in the developed more general methodology. A decision based on the new selection method will minimize biases when present in the selection model whether they originate from predisposed evaluators or unfair weighting of criteria that can skew the results.

Conclusions

Existing selection methodologies for DB teams are commonly based upon straightforward weighted averages. These methods make room for potential biases where one criterion weight or one evaluator can substantially influence the final selection decision.

Table 6. Revised Scores for All Four Teams

Technical proposal	Weight (%)	Revised scores Team 1	Team 2	Team 3	Team 4
Quality (50%)					
Experience	20	57.55	59.24	94.17	71.55
QC/QA	10	57.00	73.82	86.06	70.13
Safety	10	84.62	76.17	86.31	74.98
Performance	10	57.94	77.23	78.46	66.74
Aesthetics (20%)					
Enhance RFP	10	68.51	62.98	97.83	64.22
Involve stakeholders	10	49.24	61.90	97.17	56.63
Enhancements (15%)					
Geometric	10	5.83	45.71	92.50	67.80
Structural	5	27.67	61.28	93.00	58.38
Public relations (15%)	15	72.34	74.58	92.33	71.47
Total	100	56.06	65.88	91.17	68.00
Price (in millions)		\$178.489	\$176.938	\$233.763	\$219.000
Time (days)		392	367	437	437
Time (days × \$200,000) (in millions)		\$78.4	\$73.4	\$87.4	\$87.4
Revised score (in millions)		4.582	3.799	3.522	4.505

J. Constr. Eng. Manage., 2010, 136(8): 904-912

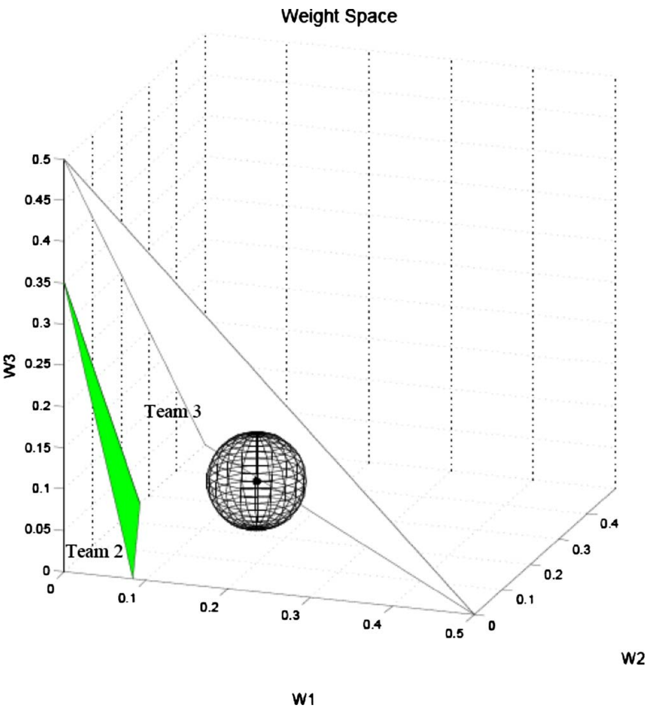

Fig. 2. Case study weight space

For this reason, there has been a need to quantify the biases in the evaluation process in order to understand and eliminate them. This is especially important in public projects because the competitors and the public need to be reassured that the selection methodology is fair. This paper presented a more statistically robust framework that can be used by practitioners to help identify and diminish biases in the evaluation in order to select the best team for the job. This would be a valuable tool in the hand of the ultimate decision makers to enable them to defend their final selection decision.

Acknowledgments

The writers wish to acknowledge the support and funding provided by the Construction and Materials Support Center (CMSC) at the University of Wisconsin-Madison in the preparation of this paper.

References

Beard, J., Loukakis, M. C., and Wundram, E. C. (2001). *Design-build: Planning through development*, McGraw-Hill, New York.

Contract Administration. (2002). *Technology and practice in Europe, international technology exchange programs*, Dept. of Transportation, Federal Highway Administration, Washington, D.C.

Dean, B. V., and Nishry, M. J. (1965). "Scoring and profitability models for evaluating and selecting engineering projects." *Oper. Res.*, 13, 550–569.

El Wardani, M. A., Messner, J. I., and Horman, M. J. (2006). "Comparing procurement methods for design/build projects." *J. Constr. Eng. Manage.*, 132(3), 230–238.

Fong, P. S.-W., and Choi, S. K.-Y. (2000). "Final contractor selection using the analytical hierarchy process." *Constr. Manage. Econom.*, 18, 547–557.

Gransberg, D. D., and Ellicott, M. A. (1997). "Best-value contracting criteria." *Cost Eng.*, 39(6), 31–34.

Gransberg, D. D., and Molenaar, K. R. (2003). "Does design/build project delivery affect the future of the public engineer?" *Proc., Transportation Research Record: TRB 2008 Annual Meeting* (CD-ROM), Transportation Research Board of the National Academies, Washington, D.C., 13–20.

Gransberg, D. D., Molenaar, K. R., Scott, S., and Smith, N. (2007). *Implementing best-value procurement in highway construction projects*, Construction Research Council, ASCE, Reston, Va.

Gransberg, D. D., and Senadheera, S. P. (1999). "Design-build contract award methods for transportation projects." *J. Transp. Eng.*, 125(6), 265–267.

Hatush, Z., and Skitmore, M. (1997). "Evaluating contractor prequalification data: Selection criteria and project success factors." *Constr. Manage. Econom.*, 15, 129–147.

Hatush, Z., and Skitmore, M. R. (1998). "Contractor selection using multicriteria utility theory: An additive model." *Build. Environ.*, 33(2–3), 105–115.

Herbsman, Z., and Ellis, R. (1992). "Multiparameter bidding system-innovation in contract administration." *J. Constr. Eng. Manage.*, 118(1), 142–150.

Martino, J. P. (1995). *Research and development project selection*, Wiley, New York.

Minnesota DOT. (2007). "Instructions to proposers." *Minnesota DOT website*, ⟨http://www.dot.state.mn.us/i35wbridge/rebuild/contractor.html⟩ (July 15, 2008).

Molenaar, K. R., and Songer, A. D. (1998). "Model for public sector design/build project selection." *J. Constr. Eng. Manage.*, 124(6), 467–479.

Palaneeswaran, E., Kumaraswamy, M., and Ng, T. (2003). "Targeting optimum value in public sector projects through 'best value'-focused contractor selection." *Construc. Architect. Manage.*, 10(6), 418–431.

Phipps, A. R. (2000). "Maine builds longest-span precast segmental bridge with unique design/build selection process." *Transportation Research Record. 1696*, Transportation Research Board, Washington, D.C., 71–75.

Postma, S. S., Carlile, F., and Roberts, J. E. (1999). "Use of best value selection process Utah Department of Transportation I-15 design-build project." *Transportation Research Record. 1654*, Transportation Research Board, Washington, D.C.

Russell, J. S. (1992). "Decision models for analysis and evaluation of construction contractors." *Constr. Manage. Econom.*, 10, 185–202.

Shane, J. S., Gransberg, D. D., Molenaar, K. R., and Gladake, J. R. (2006). "Legal challenge to a best-value procurement system." *Leadership Manage. Eng.*, 6(1), 20–25.

Singh, D., and Tiong, R. L. K. (2005). "A fuzzy decision framework for contractor selection." *J. Constr. Eng. Manage.*, 131(1), 62–70.

J. Constr. Eng. Manage., 2010, 136(8): 904-912