THE IMPORTANCE OF NON-COST CRITERIA WEIGHTING IN BEST-VALUE DESIGN-BUILD U.S. HIGHWAY PROJECTS

Maria Calahorra-Jimenez[1] Ph.D..M. ASCE, Cristina Torres-Machi, Ph.D., M. ASCE[2], Alondra Chamorro, Ph.D.[3] Luis F. Alarcón, Ph.D[4], Keith Molenaar, Ph.D., M.ASCE[5].

ABSTRACT

United States highway agencies use best-value procurement with a fixed price to select design-builders. This method enables public agencies to choose the best proposer by assessing several factors in addition to price. Theoretically, considering cost and non-cost factors in the selection enhances the probability of selecting the proposer that provides the best value for each dollar spent. However, bidding results from the last 15 years show that 80% of best-value procurements are awarded the proposer with the lowest bid. The selection seems thus to be biased towards price. This research explores the balance between cost and non-cost components in best-value procurement by identifying how weights and scores influence the selection. The goal of this analysis is to determine the ranges of weights that better balance cost and non-cost factors in the weighted criteria best-value procurement. This study characterized a first-of-a-kind dataset of 882 non-cost scores and 1,158 cost scores from 347 best-value highway projects. The study applied simulation to the weighted criteria award algorithm to explore the balance between cost and non-cost factors and derive recommendations about how to make non-cost factors more influential. The results show that weight of cost

[1] Corresponding Author. Assistant Professor, Dept. of Construction Management, Lyles College of Engineering, California State University, Fresno, 2320 E. San Ramon Ave, Fresno, CA 93740 Email:mcalahorrajimenez@mail.fresnostate.edu

[2] Assistant Professor, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Colorado, 1111 Engineering Dr. 428 UCB ,Boulder, CO 80309-0428. Email: cristina.torresmachi@colorado.edu

[3] Associate Professor, Dept. of Construction Engineering and Management, School of Engineering, Pontificia Universidad Católica de Chile, Av Vicuña Mackenna 4860, Santiago, Chile. Email: achamorro@ing.puc.cl

[4] Professor, Dept. of Construction Engineering and Management, School of Engineering, Pontificia Universidad Católica de Chile, Av Vicuña Mackenna 4860, Santiago, Chile. Email: lalarcon@ing.puc.cl

[5] Professor, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Colorado, 1111 Engineering Dr. 428 UCB ,Boulder, CO 80309-0428. Email: keith.molenaar@colorado.edu

higher or equal to 57% will result in a lowest price selection. Highways agencies should be aware of how weights and scores impact the best-value selection so that they can align these elements with their selection objectives.

Keywords: Design-build; Best-Value Procurement; Weighted Criteria; Bidding results; Highways.

INTRODUCTION

In design-build project delivery, best-value procurement is a selection method that enables public agencies to choose the proposer that provides the most advantageous offer for a particular project (AGC of America & NASFA 2008; Douglas and Michael 1997). In highway projects, the most advantageous offer relates to adding value in regards to schedule, technical merit, management options, and past performance (Molenaar and Tran 2015). This paper refers to these aspects as non-cost factors. The importance of including non-cost factors in the selection of project teams was suggested more than two decades ago. Holt et al. (1995) claimed that clients should select contractors based on the value for money rather than accepting the lowest bidder; they recommended weighting criteria related to skill, experience, and past performance. Egan (1998) reinforced this idea by arguing that procuring design and construction teams entirely based on price was one of the most significant obstacles to meeting project goals. Since then, several research have empirically demonstrated the benefits of this approach in projects' cost, schedule, and quality performance (Scheepbouwer et al. 2017).

Highway agencies have used best-value procurement in the United States (U.S.) for the last two decades (Tran et al. 2017). However, a recent study analyzing 305 projects from 18 Departments of Transportation (DOTs) procured using best-value between 2005 and 2018 has shown that 80% of the times, projects are awarded to the lowest bidder (Gaikwad 2019). Thus, U.S. design-build highway best-value procurement is biased toward price, suggesting that current practices may be missing an opportunity to balance cost and non-cost factors. Actual best-value selection results are misaligned with the concept of best-value itself, as it fundamentally differs from the lowest bid paradigm by seeking awards "on the basis of something other than the lowest cost alone" (Gransberg 2020; Ojiako et al. 2014).

Highway agencies award best-value contracts by using evaluation criteria, weights, scores, and award algorithms (Scott et al. 2006) . The evaluation criteria establish what should be measured in the proposals; this includes cost and non-cost factors. The weights represent the relevance of each criterion in the proposal's assessment, whereas the scores constitute the evaluation results for each criterion. The award algorithm refers to the formula used to combine evaluation criteria, weights, and scores to obtain an overall score.

Previous research has proposed different award algorithms (or multicriteria decision methods) to select the most suitable team for developing the contract (Alarcón and Mourgues 2002; Chen et al. 2008; Chua et al. 2001; Dobi et al. 2010; Nguyen 1986; Paek et al. 1992; San Cristóbal 2011; Scöttle et al. 2015; Seydel and Olson 1991). However, in practice, highway agencies use simplified approaches such as adjusted score (i.e., multiplying the non-cost score by the estimated project price and dividing it by the price proposal), adjusted bid (i.e., dividing bid price by the non-cost score) and weighted criteria (i.e., applying weighted sum) award algorithms. From these, the weighted criteria algorithm is the most intuitive and transparent approach because of how the evaluation criteria are weighted and scored (Molenaar and Tran 2015).

In weighted-criteria best-value procurement, weights measure the relative importance of cost and non-cost factors. Usually, these weights are based upon the relevance that each agency gives to the related criteria. General guidelines on best-value procurement leave open to the highway agency the determination of weights and scores in their best-value procurements (AASHTO 2018; U.S. Federal Goverment 2002). As a result, agencies differ in the weight ranges they apply, as explained later in this paper.

Previous research have analyzed how to determine evaluation criteria based on project characteristics (Abdelrahman et al. 2008); how weights and scores can influence the proposers' behavior (Ballesteros-pérez et al. 2016); and how subjectivity in scoring and weights might be removed using normalization and graphical models (Asmar et al. 2010). Overall, these studies have helped practitioners and academics to better understand and improve best-value procurement and have contributed to the increased use and success of this procurement method in the last years. However, none of the previous studies have analyzed whether best-value selection is balancing cost and non-cost factors and how score and weighting practices might influence this balance.

This research aims to fill this gap by addressing the question: What ranges of weights and scores can better balance cost and non-cost evaluation criteria in weighted criteria best-value procurement? To this end, this study: (1)

characterizes current scoring practices and explores their influence in the balance between cost and non-cost factors; and (2) identifies ranges of weights and scores that enable non-cost factors to be more influential in the selection.

This research constitutes a unique contribution to both scholarly literature and current practice because it presents a comprehensive analysis of 347 transportation projects procured with best-value procurement in the United States between 2002 and 2020. The results of this research will help improve existing highway agency practice by facilitating the selection of weights to use in the procurement. In summary, this research contributes to minimizing bias toward price in best-value procurement by recommending ranges of weights that help balance cost and non-cost factors in best-value selection.

The following sections include a literature review on best-value procurement, the research methodology, results, and discussion. The final section offers conclusions, contributions, recommendations for practical implications, and needs for future research.

REVIEW OF RELEVANT WORK ON BEST-VALUE PROCUREMENT

Best-value procurement aims to balance cost and non-cost factors in design-build projects. This balance provides for an evaluation of design and other non-cost factors that add value to project proposals. Thus, non-cost factors should play an essential role in the process of selecting the design-builder (DBIA 2019). However, the analysis of 15 years of best-value bidding results shows that 80% of best-value projects were awarded to the lowest bidder (FMI 2018; Gaikwad 2019). This means that best-value procurement is biased towards price and is almost operating somewhat as a low bid procurement.

Best-value procurement can be thought of as a multicriteria decision-making process that aims to answer the question of "given a set of alternatives and a set of decision criteria, what is the best alternative?" (Triantaphyllou 2000). In best-value procurement, the alternatives are the different design-builder proposals, and the decision criteria are the cost and non-cost factors that highway agencies establish to evaluate those proposals. Best-value selection requires balancing multiple factors, making it necessary to construct a model that considers the decision-maker's preferences and assessments of each evaluation criterion (Belton and Stewart 2002). The next sections summarize relevant work related to each of the components needed to obtain the overall score of proposals in best-value procurement: the award algorithm, weights, and scores.

Award algorithm

The weighted criteria is one of the award algorithms used in best-value procurement of highway projects because it is intuitive and transparent (Molenaar and Tran 2015) and has the advantage of "distinctly communicating the agency's perceived requirements for a successful proposal through the weights themselves" (AASHTO (American Association of State Highway and Transportation Officials) 2018). The weighted criteria algorithm (Equation 1) considers that having "*m*" alternatives (i.e., proposers) and "*n*" evaluation criteria, the best alternative is the one that satisfies:

$$FS = Max\ i \sum_{j=1}^{n} Wj * Sij,\ \ for\ i = 1, 2, 3, \ldots, m.\ \ \text{, with } \sum_{j=1}^{n} wj = 1. \quad \text{(Equation 1)}$$

Where: *FS* is the Final Score of the best alternative, *n* is the number of decision criteria, S_{ij}, is the score of criterion *j* in the assessment of proposal *i*, and w_j is the weight of importance of the *j* criterion.

The weighted criteria algorithm works under the implicit assumption that there exists a decision-maker's cardinal utility function, which is additive over the criteria. This means that equal *FS* can be obtained with very different proposers' performance regarding the different criteria. In other words, what is lost on one criterion is compensated by what is gained on the other (Pomerol and Barba-Romero 2000). This might lead to selections based on unbalanced criteria. To illustrate this, we can consider a best-value procurement with two evaluation criteria (i.e., cost and non-cost) and two proposers (A and B). In this particular example, let's consider that the final score of both proposers A and B is equal to 1. Based on this, both proposals have the same right to win. However, the proposer A's final score breakdown is 0.3 for cost and 0.7 for non-cost criteria, while for proposer B, it is 0.7 for cost and 0.3 for non-cost criteria. These results do not lead to a best-value selection.

Previous research on the weighted criteria algorithm has raised this issue and has proposed alternative mathematical techniques to obtain a more balanced decision (Granat et al. 2006; Pomerol and Barba-Romero 2000; Wierzbicki et al. 2000). However, these studies are theoretical in nature, and they do not address the limitations of the weighted criteria algorithm in practical approaches such as best-value procurement.

Weights

Weights should represent the relative importance of the related criterion, according to the decision-maker preferences.

Agencies adjust the weights of each evaluation criterion to reflect the needs and objectives of a particular project (Scott et al. 2006). This results in heterogeneous ranges of weights used by different public agencies. For example, a study developed in the United Kingdom found that public and private construction representatives assigned more than 60% of importance to price, with authors suggesting that assigning a maximum weight of 70% to price might help defend decision-makers from public criticism and accountability (Wong et al. 2000). In Sweden, a study analyzing 386 public bidding documents found that the weight of cost was usually set to 70% (Waara and Brochner 2006). In Australia, the Tasmanian government establishes guidelines on weighted criteria and recommends to use a weight for cost between 40% and 70% (Department of Treasury and Finance 2019).

In the U.S., general recommendations for best-value suggests a weight of cost over 50% if the cost is more important than non-cost factors (AGC of America & NASFA 2008). A report elaborated by South Carolina DOT, which summarizes design-build practices in different states, documents that South Carolina DOT typically sets a weight of cost between 50% and 70%. In contrast, Virginia DOT considers a weight for the cost of 70%, and Georgia DOT has commonly used between 50% and 80% (SCDOT 2018). Despite this, highway agencies' design-build manuals rarely recommend specific ranges to use. Some of them suggest testing the weights against different scenarios so that decision-makers can feel comfortable in case the lowest bidder is not selected (Colorado DOT 2016; LaDOT 2017).

Overall, these recommendations report the current state of best-value practice in regard to weight determination, and they show a lack of research-based criteria to set these weights. Therefore, there is a need to determine what ranges of weight for cost are more adequate to reach a best-value selection that evenly balances cost and non-cost factors.

Scores

Scores measure the level of accomplishment of each proposal towards each evaluation criterion. Public agencies have used a variety of scoring systems, from the commonly called "go/no go" to direct point assignation (Scott et al. 2006). Non-cost scores are established by a technical evaluation team. The potential bias in this evaluation might cause a significant concern (Asmar et al. 2010), leading to public mistrust and protest by bidders (Shane et al. 2006). Thus, previous research has focused on studying this potential bias—generally through case studies—and proposing methods and practices to minimize it (Asmar et al. 2010; Molenaar and Tran 2015; Tran et al. 2017). Other studies have analyzed how score rules might influence the competitiveness of bidders (Ballesteros-pérez et al. 2016), how the different types of economic scoring formulas can be categorized (Ballesteros-pérez et al. 2015), and what are the mathematical and statistical relationships between scoring parameters (González-cruz 2012).

Best-value transparency has generally been analyzed using case studies (Asmar et al. 2010; Molenaar and Tran 2015; Tran et al. 2017), while the specific analysis of scoring has been conducted in more theoretical research (Ballesteros-pérez et al. 2015, 2016; González-cruz 2012). Now that best-value has become an established procurement method in the U.S.; this study uses the opportunity to collect historical data on these procurements and analyze how best-value procurement weights and scores influence the design-build highway selection in practice.

RESEARCH METHODOLOGY

The research followed a four-step process (Fig.1). First, the authors collected and normalized historical data from 347 best-value procurements. This data was used to characterize current scoring practices using preliminary statistical analysis and distribution fitting. Based on current scoring practices, the authors analyzed the balance between cost and non-cost factors and derived recommendations on the ranges of weights and scores that enable a better balance between cost and non-cost factors. The ultimate goal of these recommendations is to inform practitioners on how non-cost factors can become more influential in the selection.

< FIGURE 1>

Data Collection and Normalization

The authors gathered cost and non-cost scores from bidding results of 347 design-build best-value highway projects procured between 2002 and 2020 by 22 DOTs. Bidding results from these projects were collected from DOTs' websites and from direct requests to DOTs' representatives. Each bid comprised two to seven cost and non-cost scores from submitted proposals. The number of scores available in each project depended on the number of firms that placed proposals for that specific project. The data set from the 347 best-value procurements included 822 non-cost scores and 1,158 cost scores.

New graph

All the scores were normalized to a common zero-to-one scale to facilitate the comparison among projects. Cost scores were normalized on the basis of the lowest bidder (Equation 2), which is a common practice in weighted criteria best-value procurement. With this normalization, the lowest bid was assigned a normalized cost score of 1, whereas the other proposers obtained a normalized score between 0 and 1 depending on how their bid compared to the lowest bid.

$$Sc_i = \frac{Lowest\ bid}{Proposer\ i\ bid} \qquad \text{(Equation 2)}$$

Where:

Sc_i is the normalized cost score in a scale 0-1, for the proposer *i*.

Lowest bid is the minimum price bid among all the proposers in the procurement.

$Proposer_i$ bid is the bid price of proposer *i*.

Non-cost scores were also normalized to a zero-to-one scale. The initial non-cost scores had different scales (e.g., 1-100, 1-1,000, 1-1,200) depending on the scale used in the procurement. The authors normalized each bid's non-cost scores to a common 0-1 range using Equation 3, which conserves the proportionally between scales.

$$Snc_i = \frac{Initial\ nc\ score_i}{Max\ initial\ nc\ score} \qquad \text{(Equation 3)}$$

Where:

Snc_i is the normalized non-cost score in a scale 0-1 for proposer *i*.

Initial nc $score_i$ is the non-cost score of proposer *i* based on the initial procurement scale.

Max initial ns score is the maximum value of non-cost scores based on the initial procurement scale.

Characterization of Current Scoring Practices

The characterization of current scoring practices consisted of a preliminary analysis of normalized scores (including basic descriptive analysis, outlier detection, and the analysis of scores variability) and the fitting of probability functions that best represent the score dataset.

Preliminary Analysis

A descriptive analysis was conducted to determine the main statistics for cost and non-cost scores and their distribution in histogram diagrams. Following this, a variability analysis was performed to ensure that the scoring data were homogeneous in terms of the project scope and geographic distribution. The projects considered in the research had varying scopes, including bridges, highways, and interchanges. Therefore, a variability analysis was performed to determine whether the project scope impacted scoring. A similar analysis was performed to determine the potential impact of the geographic distribution of data. This analysis was necessary because some DOTs had a significantly larger experience in best-value than other DOTs and, therefore, contributed to a larger set of data. The variability analysis sought to identify potential differences in scoring practices among states. Both analysis (variability based on project scope and state) were based on the Mann-Whitney U test (Conover 1980).

Finally, an analysis of outliers was conducted to quantify, characterize, and determine how to treat this type of data. A score data was identified as an outlier if it was outside the range (Q1-1.5IQR, Q3+1.5IQR), where Q1 is the first quartile, Q3, the third quartile and IQR the interquartile range. Once the outliers were identified, the authors analyzed the potential reasons why each of these data points were outliers and derived conclusions on whether they should be removed or not from the analysis on a case-by-case basis.

Goodness-of-fit Analysis

The authors developed a statistical analysis to determine the probability distributions that best fit the cost and non-cost scores. These probability distributions were used in the simulation process to characterize current practices in scoring and simulate their impact on the final evaluation. To find the probabilistic distributions, goodness-of-fit techniques were used to measure the fitness of the sample with a set of hypothesis distributions (D'Agostino and Stephens 1986). Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) were used to identify the distribution providing the best fit for the data. AIC (Akaike 1974) is a technique based on in-sample fit to estimate the likelihood of a model to predict future values. BIC (Stone 1979) is another criterion for model selection that measures the trade-off between the model complexity and fit. These metrics do not have physical meaning, except in relative terms, with the lower parameter being indicative of a better fit (Yoe 2019). The probability distributions selected were those with the lowest parameters for both AIC and BIC methods.

Analysis of the Balance between Cost and Non-Cost Factors

The cost and non-cost probability distributions resulting from the previous analysis were used to analyze the weighted criteria algorithm (Equation 1) under different weighting scenarios. In each scenario of weight, the balance between cost and non-cost factors was assessed.

Metric to measure balance

To understand how cost and non-cost factors impact the final score in weighted criteria algorithm, the research considered a two-component algorithm comprising cost and non-cost criteria (Equation 4). The balance between cost and non-cost factors was measured in terms of the ratio *RC,* which represents the proportion of cost over non-cost factors in the final score (Equation 5).

$$FS = Wc * Sc + Wnc * Snc, \text{ with } Wc + Wnc = 1 \qquad \text{(Equation 4)}$$

$$\mathrm{RC} = \frac{Wc * Sc}{Wnc * Snc} \qquad \text{(Equation 5)}$$

FS is the final score; Wc and Wnc represent the weights of cost and non-cost factors and, Sc and Snc account for the score of cost and non-cost factors, respectively. The cost component ($Wc*Sc$) relates to the weight of cost multiplied by the cost score. Similarly, the non-cost component ($Wnc*Snc$) is obtained by multiplying the weight and the score assigned to the non-cost factors. RC constitutes the ratio between the cost and the non-cost component.

The ratio between cost and non-cost components is relevant because high or low RC values are indicators of unbalanced selections (Fig.2). High RC values represent cases where the cost component is notably larger than the non-cost component, leading thus, to a cost-driven selection. On the contrary, low RC values imply that the non-cost component has more significant importance than cost, which may lead to a non-cost-driven selection. Best-value procurement aims to select the best contractor on the basis of a balanced evaluation of cost and non-cost factors. Therefore, extreme RC values should be avoided to ensure a best-value selection.

< FIGURE 2 >

Monte Carlo Simulation

Each procurement might have different values for weights and scores, resulting in a balance that can be deterministically calculated. However, the analysis of particular cases does not enable researchers to find a fairly accurate estimate of that balance of cost and non-cost in the best-value practice as a whole. For this reason, this research used Monte Carlo simulation. By using this technique, it is possible to estimate a deterministic quantity by using a large and random sample (Brandimarte 2014). In this study, two probability distributions built upon empirical score data were considered to simulate final project scores and find the balance between cost and non-cost factors in different weighting scenarios.

According to Johnson (2013), simulation is a “way of forming an educated guess about the most likely outcomes or the range of possibilities.” Through Monte Carlo simulation, researchers can obtain enough large set of results that enables them to make statistical inferences (Kroese et al. 2014). In this research, the authors used Monte Carlo simulation to replicate a large number of weighted criteria best-value procurements using different sets of weights and scores. In each iteration, the relative contribution of the cost component, and the non-cost component (i.e., RC) was analyzed.

The authors determined 0.3 and 0.7 as the extreme values for the weight of cost. This range contains 70% of the dataset weighted criteria cases. Further, this range consider a variation of ± 20% in regards 50%. Fifty percent (50%) is the limit suggested by AGC of America % NASFA (2008) when cost is more important than non-cost factors.

Within this range, the authors performed two analyses. The first one, considering equal weights ($Wc = Wt = 0.5$), aimed to better understand the impact of current scoring practices in the overall evaluation. The second analysis considered 41 scenarios with weights varying in centesimal increments (e.g., in the scenario 1, Wc=0.3; scenario 2, Wc=0.31: scenario 3, Wc=0.32; etc). For each of these scenarios, the weight of non-cost factors (Wnc) was determined by considering that $Wc+Wt$=1 (Equation 4). In each scenario, the simulation run iterations in which the cost and non-cost scores were obtained from the probability distributions. The Monte Carlo simulations were performed using @Risk software, considering a seed to guarantee replicability. To reach validity in the results, each weight scenario comprised 10,000 iterations to ensure the convergence of the output mean and standard deviation with a 95% confidence level.

Recommendations of Ranges of Weights and Scores

Finally, the simulation results were analyzed to derive recommendations of the ranges of weights and scores that should be used to make non-cost factors more influential in best-value selection.

RESULTS

Characterization of Current Scoring Practice

Preliminary Analysis

The original data distribution is characterized by descriptive statistics (Table 1) and histograms (Fig. 3 and Fig. 4).. When analyzing these results it is important to remember that all the scores were normalized to a common zero-to-one scale (using Equations 2 and 3) to facilitate the comparison among projects. Results from the preliminary descriptive analysis show that cost scores are slightly more skewed toward “1” than non-cost values (median 0.879 vs. 0.865 in Table 1). The data spread is higher in cost scores (standard deviation of 0.103 vs. 0.079).

< TABLE 1>

To understand the shape and the spread of the sample, the data are displayed using histograms in the form of relative frequency graphs. The cost score data pattern is skewed to the left, with 90% of the cost scores ranging from 0.652 and 0.987. Fig.3 shows a gradual variation of the data frequency from the peak (0.88-0.93) to the minimum and maximum values (0.53 and 0.99).

< FIGURE 3 >

Non-cost scores are slightly skewed to the left, with 90% of the non-cost scores ranging from 0.610 and 0.958. In contrast with the cost scores, Fig.4 shows a sharp variation of the data frequency from the peak (between 0.77 and 0.93) to both high and low ends. The values lower than 0.77 and higher than 93 are much unlikely.

< FIGURE 4 >

The variability analysis shows that neither project type nor project location have a significant impact on scoring. The results of the Mann Whitney U test (p-value > 0.2) show that there is no statistically significant difference among project scopes for both cost and non-cost scores. Concerning the project’s geographic distribution, Florida´s projects constitute 35% of the whole sample. Given this, it was tested whether there was a statistically significant difference in the scores of Florida´s projects and the overall sample. The results of the Mann Whitney U test (p-value 0.154) show that there is not a statistically significant difference. Therefore, the data set is considered consistent and was not divided based on project type nor location.

Finally, the authors identify the outliers in the score sample. This analysis determines that cost scores have 16 outliers, accounting for roughly 2% of the cost scores' sample. These data points represent bids with prices 70% higher or more than the lowest bid (i.e., cost scores lower than 0.579). In this research, the outliers in the cost score sample correspond to six (6) projects, with three (3) of them having two (2) outliers. Procurements with two (2) or more outliers indicate that more than 60% of the proposers bid outside the expected range. This suggests a very high variance in bid prices and may result from specific project-case circumstances that do not represent standard practices. This might happen, for example, when one company bid with a very low and unrealistic price aiming to win the contract. As a result, other companies score very low (because they are costly) as compared with the lowest bidder. The authors considered that having two (2) or more outliers do not reflect the general scoring trend. Thus, the scores of three (3) projects containing six (6) outliers were removed.

The non-cost scores contain 52 outliers, accounting for 6% of the sample. All the outliers have values lower than 0.66. Outliers in non-cost scores might correspond to specific project-case circumstances. It may be the result of vague Request for Proposals (RFP) that lead most of the proposers to not adequately prepare their proposals. In this study, non-cost scores outliers correspond to 17 projects from six (6) DOTs. Twelve (12) of these projects have more than two (2) outliers, representing a total of 34 outliers. One DOT contributed the most to the outlier set, with five (5) projects having 17 outliers. In this particular case, projects were delivered between 2014 and 2016, just when this DOT began using best-value procurement. The second DOT with significant contribution provided three (3) projects with a total of 7 outliers. These projects were delivered between 2010 and 2012, also in the early years of using best-value procurement. The remaining four DOTs provided only one project each. These projects were delivered at different times (2007, 2012, 2015, 2018), suggesting that each specific case's circumstances might explain the outliers.

Overall, these cases do not correspond to the general scoring trend, which is what this research aims to simulate. Therefore, the scores associated with the 12 projects that contained the 34 outliers were removed for further analysis.

Goodness-of-fit analysis

The Beta distribution is the probability distribution that better fit cost scores. They had thus the lowest values for both AIC and BIC parameters (Table 2).

< TABLE 2 >

The Beta distribution is commonly used to describe variability over a limited range, being naturally defined over 0 and 1 (Yoe 2019). The Beta distribution is widely known as the foundation of the Program Evaluation and Review Technique (PERT) method. The PERT method is usually adopted to model task duration in construction management by using three values, the most optimistic (shorter), the most pessimistic (longest), and the most likely (mode) (Damnjanovic and Reinschmidt 2020). Cost scores could indeed be characterized in this way by considering a maximum value of 1, a minimum value of approximately 0.65, and a most likely range between 0.87-0.93.

Non-cost scores were fitted to a Gumbel distribution. The Gumbel distribution is a limiting extreme value distribution that serves to model the maximum and minimum values of any set of data (Gumbel 1955). The Gumbel distribution is used to model extreme events as well as construction design elements (Mun 2002). The non-cost scores distribute at the high end of the evaluation scale, with 70% of the data between 0.77 and 0.93. It seems thus reasonable that non-cost scores are well suited for the Gumbel distribution because they do not normally vary around one value.

Analysis of the Balance between cost and non-cost factors

The balance between cost and non-cost factors was measured in terms of the RC ratio, which represents the proportion of cost over non-cost factors in the final score (Equation 5). RC values were obtained using Monte Carlo simulation and the Beta and Gumbel distributions to characterize current scoring practices and explore different weighting scenarios. For the simulation, both distributions were truncated in the maximum and minimum value of "1" and "0", respectively.

RC with Equal Weights

Intuitively, when setting equal weights to both factors, the decision-maker would expect a balanced contribution of cost and non-cost factors in the overall score. However, the simulation showed that, based on current scoring practices, overall scores did not follow this intuition. Although the weights were equal, the cost component had a more significant contribution to the overall score than the non-cost component. This is shown in the probability density graph depicted in Fig.5, where RC is higher than "1" in 68.6% of the cases.

< FIGURE 5>

If both components were to contribute evenly to the overall score, the relative frequency graph would be symmetric and centered in 1 (implying that the cost and non-cost component have equal relative importance). However, the results show a relative frequency skewed to the right, meaning that when assigning equal weights, the cost component has a larger contribution to the overall score compared to the non-cost component in 68.6% of the cases. These results are explained because the cost scores are statistically higher than the non-cost scores (Fig.3 and Fig.4). Ultimately, this led to a counterintuitive result in which equal weights do not result in the equal importance of cost and non-cost factors in the overall score. Therefore, there is a need to better understand how current scoring practices are impacting the relative importance of cost and non-cost factors in best-value procurement. To do so, the next section analyzes the impact of different weighting scenarios on the relative importance of cost and non-cost components.

RC with Different Weighting Scenarios

Fig.6 synthesizes the results from the simulation of 41 weighting scenarios. The X-axis represents the weight of cost considered in each scenario (*Wc*), whereas the Y-axis represents the proportion of cost over non-cost factors in the overall score (RC). For each value of weight of cost, a Monte Carlo simulation with 10,000 iterations is run, and the values of RC for the median and 5^{th} and 95^{th} percentile are recorded. These values, plotted in Fig.6 with solid black lines, account for 90% of the cases in each scenario of weight. The grey dashed line represents the ratio defined by the weights of cost (*Wc*) and non-cost factors (*Wnc*). In other words, it is the contribution of the weights to the RC ratio (*Wc/Wnc*). If RC is higher than the value defined by the grey dashed line, this means that the cost score is higher than the non-cost score. The median (percentile 50%) is above this line; therefore, in 50% of the procurements, cost scores are higher than non-cost scores.

< FIGURE 6 >

The area defined above $RC = 1$ shows a cost-driven selection, where the cost component is higher than the non-cost component. On the contrary, the area below $RC = 1$ represents the non-cost-driven selection, where the non-cost component is higher than the cost component. Given this consideration and the results obtained in the simulation,

Fig.6 shows that weights of cost lower than 0.43 result in a non-cost-driven selection. On the contrary, weights of cost higher than 0.57 lead to a cost-driven selection. In the range of weights defined between 0.43 and 0.57, the result of the selection can be either cost or non-cost driven.

Another interesting result relies on the increasing distance between percentiles 5th and 95th as the weight of cost increases. This relates to the relative effect of weights and scores on the RC ratio. Low weights of cost (Wc), result in low ratios between the weights (e.g., for Wc = 0.3, Wc/Wnc is equal to 0.3/07 = 0.43; while for Wc = 0.7, Wc/Wnc is equal to 2.33). This trend is represented by the grey dashed line in Fig.6. As the weight of cost increases, the ratio of the scores (Sc/Snc) is therefore multiplied by a higher number, resulting thus in more spread RC values and larger distance between the 5th and 95th percentile lines.

DISCUSSION

This research aimed to find the ranges of weights that lead to a better balance between cost and non-cost factors in the weighted criteria best-value procurement. By analyzing historical cost and non-cost scores under different weighting scenarios, the study provided insight into what ranges might lead to a better balance in best-value selection.

Furthermore, this research raises awareness about how the trade-off between cost and non-cost criteria can vary depending on the award algorithm, the weights and the scores used. Specifically, the weighted criteria algorithm (because of the nature of the formula) might lead to selections based on unbalance criteria. Previous research proposed mathematical techniques to obtain more balance selections when using the weighted criteria algorithm (Granat et al. 2006; Pomerol and Barba-Romero 2000; Wierzbicki et al. 2000). This research contributes to scholarly research on weighted criteria algorithm by proposing a range of weights that limit the relative variation of cost and non-cost criteria. Thus, this research addresses the weighted criteria limitation under the practical approach of best-value procurement selection.

Weights

The analysis showed that, under current scoring practices, weights lower than 0.43 and higher than 0.57 do not enable highway agencies to make a best-value selection, as the selection is skewed toward either non-cost or cost factors. A range of weight of costs between 0.43 and 0.57 allows decision-makers to have chances of having both cost and non-cost-driven selections. In this range, the scores can determine whether the non-cost component is higher or lower than the cost component. Further, this range of weights ensures that in 90% of the cases the RC is between 0.6 and 1.8. This means that the proportion of the cost component regarding the non-cost component is limited within a range that minimizes selections based on unbalanced criteria

It is relevant to note that the proposed range of weight of cost between 0.43 and 0.57 is due to the existing scoring tendencies (represented by Fig. 3 and Fig.4). This range could be wider for a specific highway agency if this agency followed a wider pattern in the scoring of the proposals, as suggested in the following section.

Overall, the range of weights that this research proposes helps to minimize unbalance selections in the weighted sum algorithm when applied in best-value procurement. Further, previous studies and the current state-of-the-practice showed a lack of specific criteria to determine the weights to use in weighted criteria best-value procurement (AGC of America & NASFA 2008; Colorado DOT 2016; LaDOT 2017; SCDOT 2018). This research contributes to this current state of practice by proposing the use of weights of cost between 0.43 and 0.57 in order to reach an adequate balance between cost and non-cost factors in the selection. These research-based recommendations might help highway agencies to overcome the fear suggested by Wong et al. (2000) about using weight for cost lower than 70%.

ScoresAs well as weights, the scores might play an essential role in characterizing the best-value selection. Previous research analyzed best-value scoring using case studies (Asmar et al. 2010; Molenaar and Tran 2015; Tran et al. 2017), or theoretical approaches (Ballesteros-pérez et al. 2015, 2016; González-cruz 2012). This research contributes to scholarly research on best-value scoring by collecting and analyzing a dataset of 882 non-cost scores and 1,158 cost scores from 347 best-value highway projects.

The cost and non-cost data distributions (represented by Fig. 3 and Fig.4) served to obtain the best-value range of weight of cost between 0.43 and 0.57. This range could vary for a specific highway agency if this agency had another scoring pattern. In this regard, this research proposes the following recommendations for non-cost and cost scores.

Non-cost scores.

Widening the range of non-cost scores can help to make the non-cost component more influential. Under a stochastic approach (such as the one developed in this research), having a more spread non-cost score distribution (similar to the cost scores' one) would lead to a balance between cost and non-cost factors comparable to the one established by the weights. In other words, taking the example of equal weights shown in Fig.5, similar spread in cost and non-cost distributions would lead to a more symmetric and centered in “1” RC distribution.

Under a deterministic approach, considering a wide range of non-cost scores in each procurement would enable highway agencies to make a more meaningful differentiation between the technical proposals. In other words, if all the proposers score equally in the technical evaluation (meaning a narrow range of non-cost scores so that there is no “technical” distinction among proposers), the differentiator would not be the non-cost component, but solely the cost. Thus, this research recommends widening the non-cost scores range in best-value technical evaluation in order to make non-cost factors more influential.

Cost-scores

Expectations on cost scores dispersion can help highway agencies to decide the weight of cost to use in each procurement. The results of this research are based upon a historic cost dispersion. However, agencies might expect a very “tight” or a very “wide” range of price proposals depending upon each project’s scope, risks, or innovation.

Expecting ranges of cost scores that are narrow, and close to “1” implies that all the bids are close to the lowest bidder. Having all cost scores on the upper side of the evaluation scale makes it more likely that non-cost scores are lower than cost scores. In this case, highway agencies might wish to weight up non-cost factors by using a lower weight of cost. On the contrary, expecting a wide range of cost scores suggest a more likely trade-off between cost and non-cost factors in each proposer’s evaluation. In this case, highway agencies should consider the recommendations given for both weights and non-cost scores in previous sections.

CONCLUSIONS

The goal of balancing cost and non-cost criteria in best-value selection is not being realized. Highway agencies use non-cost criteria to evaluate and select design-builders. However, in more than 80% of the cases in our dataset, the best-value selection award the contract to the lowest bidder. This evidence shows a bias toward price of best-value selection; in other words, a lack of balance between cost and non-cost factors. This research aimed to solve this problem by addressing the following research question: What ranges of weights and scores can better balance cost and non-cost evaluation criteria in the weighted criteria best-value procurement?

The findings showed that the weight of cost and the ranges of scores used in the evaluation play an essential role in having cost or non-cost-driven selections. Indeed, weights of cost higher than 57% always lead to a cost-driven selection; that is, a low bid selection. A weight of cost ranging between 43% and 57% strengthens the best-value selection by enabling highway agencies to reverse the driver of the selection depending on the difference between cost and non-cost scores. In this range of weight, the selection might be cost or non-cost-driven. Further, by using this range of weights, highways agencies will prevent selections based on unbalanced criteria. This is because this range of weights limits the cost component to be bounded into 1.8-0.6 times the non-cost component.

Weights between 43% and 57% not only enable highway agencies to balance cost and non-cost factors in the selection. Also, they send the message that both cost and non-cost factors are important. This might minimize the proposers' tendency to cut bid prices in order to be the lowest bidder. Instead, by using this range of weights, the idea of providing "the best-value for dollar spent" is encouraged.

It is relevant to note that the proposed range of weight of cost between 43% and 57% is based on the existing scoring tendency. The historical scoring pattern used in this research showed a skewness toward "1" for both cost and non-cost scores and a more widespread distribution of cost scores as compared with non-cost scores. This tendency leads the cost component in the evaluation to be more influential than the non-cost component.

However, the proposed range of weight of cost could vary if the scoring trends were different. Highway agencies could consider a wider range of weights of cost by using a wider pattern in the scoring of the proposals. Specifically, non-cost factors could be more influential if the proposals' evaluation led to non-cost scores within a range wide enough to enable the differentiation among the proposers. Further, another aspect to consider would be the cost score dispersion. Non-cost factors could be more influential if highway agencies considered the effect of potential economic bid dispersion when selecting the weight of cost. Low dispersion of cost scores suggests more probability of having cost scores higher than non-cost scores. Thus, in these cases, highway agencies might adopt a lower weight of cost in order to balance cost and non-cost factors.

In summary, this research constitutes a unique contribution to both scholarly literature and current practice by recommending the weight of cost range that should be used to properly balance cost and non-cost factors in the weighted criteria best-value procurement. This recommendation is derived from the analysis of 347 highway projects using best-value procurement over the last two decades, reflecting thus existing trends in current practice. The use of this range of weights will contribute to minimizing bias toward cost in best-value procurement.

This research establishes the first step to minimize the bias toward cost of best-value practice, which will make more influential non-cost factors in the selection and, in turn, will increase the likelihood of achieving project goals. This research, however, did not address all the elements that might influence best-value results and the balance between cost and non-cost factors. These topics are thus suggested for future research. This includes the analysis of different award algorithms and scoring systems and how they might impact the balance between cost and non-cost factors. Further, the evaluation criteria selected to represent the non-cost factors might also influence the scoring trend and, therefore, the balance between cost and non-cost factors. Another practical and future research will consider case studies to evaluate the range of weights proposed in this research.

DATA AVAILABILITY STATEMENT

Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

ACKNOWLEDGMENTS

The writers wish to acknowledge all the state highway agency personnel who supported this research. Maria Calahorra-Jimenez would like to thank the economic support given by the Vice-Rector's Office of Research (VRI) of the Pontificia Universidad Católica de Chile.

REFERENCES

AASHTO (American Association of State Highway and Transportation Officials). (2018). *AASHTO Guide for Design-Build Procurement*. AASHTO, Washington D.C.

Abdelrahman, M., Zayed, T., and Elyamany, A. (2008). "Best-Value Model Based on Project Specific Characteristics." *Journal of Construction Engineering and Management*, 134(3), 179–188.

AGC of America & NASFA. (2008). *Best Practices for use of Best Value Selections*.

Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Alarcón, L. F., and Mourgues, C. (2002). "Performance Modeling for Contractor Selection." *Journal of Management* *in Engineering*, 18(2), 52–60.

Asmar, M. El, Lotfallah, W., Whited, G., Asce, M., Hanna, A. S., and Asce, F. (2010). “Quantitative Methods for Design-Build Team Selection.” *Journa of Construction Engineering and Management*, 136(August), 904–912.

Ballesteros-pérez, P., Skitmore, M., Pellicer, E., Asce, M., and Zhang, X. (2016). “Scoring Rules and Competitive Behavior in Best-Value Construction Auctions.” *Journal of Construction Engineering and Management*, 142(9), 1–14.

Ballesteros-pérez, P., Skitmore, M., Pellicer, E., Carmen, M., Skitmore, M., and Ballesteros-pe, P. (2015). “Scoring rules and abnormally low bids criteria in construction tenders : a taxonomic review Scoring rules and abnormally low bids criteria in construction tenders : a taxonomic review.” *Construction Management and Economics*, Routledge, 33(4), 259–278.

Belton, V., and Stewart, T. J. (2002). *Multiple Criteria Decision Analysis. An integrated Approach*. Springer, Dordrecht.

Brandimarte, P. (2014). *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*. Wiley.

Chen, Z., Li, H., Ross, A., Khalfan, M. M. A., and Kong, S. C. W. (2008). “Knowledge-Driven ANP Approach to Vendors Evaluation for Sustainable Construction.” *Journal for Construction Engineering and Management*, 134(12), 928–941.

Chua, D. K. H., Li, D. Z., and Chan, W. T. (2001). “Case-based reasoning approach in bid decision making.” *Journal for Construction Engineering and Management*, 127(1), 35–45.

Colorado DOT. (2016). “Design-Build Manual.”

Conover, W. J. (1980). *Practical Nonparametric Statistics*. John Wiley & Sons.

D’Agostino, R. B., and Stephens, M. A. (1986). *Goodness-of-Fit Techniques.* Marcel Dekker, New York.

Damnjanovic, I., and Reinschmidt, K. (2020). *Data Analytics for Engineering and Construction Project Risk*. Springer Nature Switzerland AG 2020.

DBIA. (2019). “Position Statement Principles of Best Value Selection Principles of Best Value Selection.”

Department of Treasury and Finance. (2019). *Guidelines on Tender Evaluation using Weighted Criteria for Building Works and Services*. *Evaluation*, Hobart.

Dobi, K., Gugić, J., Kancijan, D., Education, T., and Jagera, L. (2010). “AHP As a Decision Support Tool in the Multicriteria Evaluation of Bids in Public Procurement.” *Proceedings of the ITI 2010, 32nd International Conference on Information Technology Interfaces*, IEEE, 447–452.

Douglas, D., and Michael, A. (1997). “Best-value contracting criteria.” *Cost Engineering*, 39(6), 31–34.

Egan, J. (1998). “Rethinking Construction the Report of the Construction Task Force.” *Construction*, 38.

FMI. (2018). “Design-Build Utilization. Combined Market Study.”

Gaikwad, S. V. (2019). “Challenges in engineering estimates for best-value design-build projects: an analysis of bid dispersion in US highway projects.” University ofColorado Boulder.

González-cruz, M. C. (2012). “Mathematical relationships between scoring parameters in capped tendering.” *International Journal of Project Management*, Elsevier Ltd, 30(7), 850–862.

Granat, Makowski, and Wierzbicki. (2006). “Hierarchical reference approach to multi-criteria analysis of discrete alternatives.” *CSM’ 06: 20th Workshop on Methodologies and Tools for Complex System Modeling and Integrated Policy Assessment. IIASA, Laxenburg, Austria*.

Gransberg, D. D. (2020). “Does low bid award facilitate wrongdoing? US implications of quebec’s charbonneau commission report.” *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 12(1), 1–6.

Gumbel, E. J. (1955). *Statistical Theory of Extreme Values and Some Practical Applications*. *The Mathematical Gazette*.

Holt, G. D., Olomolaiye, P., and Harris, F. C. (1995). “A Review of Contractor Selection Practice in the U . K . Construction Industry.” *Building and Environment*, 30(4), 553–561.

Johnson, P. E. (2013). “Monte Carlo Analysis in Academic Research.” (or MCMC), 1–67.

Kroese, D. P., Brereton, T., Taimre, T., and Botev, Z. I. (2014). “Why the Monte Carlo method is so important today USES OF THE MCM.” *WIREs Comput Stat*, 6, 386–392.

LaDOT. (2017). “Louisiana Department of Transportation and Developement.”

Molenaar, K. R., and Tran, D. (2015). *NCHRP Synthesis 471. Practices for Developing Transparent Best Value Selection Procedures*. Transportation Research Board of the National Academies, Washington, D.C.

Mun, J. (2002). *Advanced Analytical Models*. John Wiley & Sons.

Nguyen, V. U. (1986). “Tender Evaluation by Fuzzy Sets.” *Journal for Construction Engineering and Management*, 111(3), 231–243.

Ojiako, U., Park, J. H., Williams, T., Marshall, A., and Chipulu, M. (2014). “What is best value in public sector building construction?” *Proceedings of Institution of Civil Engineers: Management, Procurement and Law*, 167(5), 209–219.

Paek, B. J. H., Lee, Y. W., and Napier, T. R. (1992). “Selection of design/build proposal using f u z z y - l o g i c system.” *Journal for Construction Engineering and Management*, 118(2), 303–317.

Pomerol, J.-C., and Barba-Romero, S. (2000). *Multicriterion Decision in Management: Principles and Practice*. Springer, New York.

San Cristóbal, J. R. (2011). “Contractor Selection Using Multicriteria Decision-Making Methods.” *Journal of*

Construction Engineering and Management, 138(6), 751–758.

SCDOT. (2018). *2018 Design-Build Peer Exchange*.

Scheepbouwer, E., Gransberg, D. D., and Lopez del Puerto, C. (2017). "Construction engineering management culture shift: Is the lowest tender offer dead?" *Frontiers of Engineering Management*, 4(1), 49–57.

Scott, S., Molenaar, K. R., Gransberg, D. D., and Smith, N. C. (2006). *NCHRP Report 561. Best-Value Procurement Methods for Highway Construction Projects*. Transportation Research Board of the National Academies, Washington, D.C.

Scöttle, A., Arroyo, P., and Bade, M. (2015). "Comparing Three Methods in the Tendering Procedure to Select the Project Team." *23rd Annual Conference of the International Group for Lean Construction*, 267–276.

Seydel, J., and Olson, D. L. (1991). "Bids considering multiple criteria." *Journal for Construction Engineering and Management*, 116(4), 609–623.

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

Stone, M. (1979). "Comments on Model Selection Criteria of Akaike and Schwarz." *Journal of the Royal Statistical Society*, 41(2), 276–278.

Tran, D. Q., Molenaar, K. R., and Kolli, B. (2017). "Implementation of best-value procurement for highway design and construction in the USA."

Triantaphyllou, E. (2000). *Multi-criteria decision making methods: a comparative study*. Springer.

US Federal Goverment. (2002). *Code of Federal Regulations. Part 636. DB Contracting*. 212–228.

Waara, F., and Brochner, J. (2006). "Price and Nonprice Criteria for Contractor Selection." *Journal of Construction Engineering and Management*, 132(August), 797–804.

Wierzbicki, A., Paczynski, J., and Makowsky, M. (2000). *Modeling tools*.

Wong, C. H., Holt, G. D., and Cooper, P. A. (2000). "Lowest price or value? Investigation of UK construction clients' tender selection process." *Construction Management and Economics*, 18(7), 767–774.

Yoe, C. (2019). "Choosing a probability distribution." *Principles of Risk Analysis. Decision Making Under Uncertainty*, Taylor & Francis Group,LLC, Boca Raton, 319–361.

TABLES

Table 1 Descriptive statistics

	Mean	Mode	Median	Variance	Std. Deviation
Cost Scores [Sc]	0.857	0.896	0.879	0.011	0.103
Non-cost Scores [Snc]	0.853	0.820	0.865	0.013	0.079

Table 2 Cost and Non-Cost Scores. Goodness-of-fit-parameters

	Cost Scores		Non-Cost Scores	
Distributions/parameters	**AIC**	**BIC**	**AIC**	**BIC**
Beta	**-1,612**	**-1,593**	n/a	n/a

Gumbel	-1,546	-1,537	**-1,930**	**-1,921**
Logistic	-1,407	-1,397	-1,858	-1,849
Normal	-1,370	-1,361	-1,727	-1,718
Laplace	-1,347	-1,338	-1,860	-1,851
Triangular	-940	-926	-1,222	-1,209

620