STRUCTURED APPROACH FOR BEST-VALUE EVALUATION CRITERIA: US DESIGN-BUILD HIGHWAY PROCUREMENT

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

ABSTRACT

In best-value procurement, current practice shows that cost is frequently more influential than non-cost factors, and consequently, the lowest bidder is chosen in most of the cases; thus, a best-value selection is not achieved. Design-builders cannot offer the best-value in their proposals if evaluation criteria do not show precisely what constitutes best-value and how best-value is scored. Thus, the aim of this research is twofold: first, to identify how highway agencies articulate evaluation criteria; and second, to propose a structured approach that enhances current practice on writing evaluation criteria. Through the lens of decision analysis, the researchers conducted a content analysis on 540 evaluation criteria included in 98 requests for proposals (RFPs) from 21 states across the United States (US). The study showed that 43% of evaluation criteria were generic, 53% used a generic constructed scale, and 4% assigned points or levels directly. These three groups represent different levels of specificity in writing evaluation criteria. Building upon these levels and on decision analysis theory principles, this research proposes a structured approach to support highway agencies in the process of crafting evaluation criteria. More precise and specific evaluation criteria

---

[1] Corresponding Author. Ph.D. candidate, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Colorado, 1111 Engineering Dr. 428 UCB, Boulder, CO 80309-0428. Ph.D. candidate, Dept. of Construction Engineering and Management, Univ. Católica de Chile, Av Vicuña Mackenna 4860, Santiago, Chile. Email: maca9889@colorado.edu

[2] 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

[3] 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

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

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

will enhance the proposals' ability to offer the best-value, which, in turn, will enhance the best-value selection process as a whole.

Keywords: decision analysis; design-build; best-value procurement; evaluation criteria; highways.

INTRODUCTION

Best-value procurement is a decision-making process to choose between competitive proposals. According to Molenaar & Tran (2015), best-value procurement is a process to select the most advantageous proposal by evaluating other factors in addition to price. These factors should be determined in alignment with the project's goals and assessed through evaluation criteria (Scott et al. 2006). In best-value procurement for design-build highway projects, the project's goals should translate into objectives that highway agencies will use to select the proposal that provides the best-value to the agency. To measure the level of accomplishment of the objectives, highway agencies ask design-builders proposers to submit different pieces of information, called requirements. Agencies assess these requirements through evaluation criteria to determine the level of accomplishment of the objectives. The level of achievement of the objectives is the basis of the design-builders' selection. This selection should be logically consistent with (1) the information obtained from the design-builders, and (2) the preferences that decision-makers establish in their evaluation criteria (Howard 1988). Evaluation criteria measure to what extent the design-builders accomplish the selection objectives by assessing the requirements. Award algorithms consider decision-makers' preferences by combining evaluation criteria's scores and their related weights.

Decision analysis is "a theoretical paradigm for decision making and a body of practical experience for using this paradigm to illuminate the decision problem to the decision-maker" (Howard 1980). Decision analysis concerns the capacity to formulate adequate evaluation criteria. These criteria enable decision-makers to obtain meaningful information from the design-builders and, therefore, to make valuable comparisons among them (Keeney and Gregory 2005; US Federal Government 2002). Established decision analysis theory states that evaluation criteria should be comprehensive, direct, unambiguous, and understandable (Belton and Stewart 2002; Keeney and Gregory 2005; Keeney and Raiffa 1976). This means that evaluation criteria need to cover the full range of potential variability in the requirements, or in other words, they need to be comprehensive. It also needs to establish a direct relationship between its levels, and the requirements asked to measure the objectives. Further, evaluation criteria should not be vague or imprecise in their definition but rather be unambiguous. Finally, the result of the evaluation criteria assessment should readily be understood and clearly communicated.

In US design-build best-value procurement, highway agencies consider non-cost factors (Choi et al. 2020; Papajohn et al. 2019) by including a wide variety of evaluation criteria. However, previous research has pointed out that the selection of design-builders is heavily skewed toward price (FMI 2018; Gaikwad 2019). The design-build utilization study by FMI (2018) determined that, despite using non-cost selection criteria, highway agencies make the majority of their selection based on price. Further, reviews of the bidding results of 305 DB best-value highway projects in the US also reveals that 80% of the projects were awarded to the lowest bidder on a sample of projects from 15 states between 2005 and 2018 (Gaikwad 2019). Thus, although highway agencies consider non-cost criteria in the evaluation, most of these criteria are *not* influential in the decision. This evidence identifies an opportunity to enhance the procurement process to ensure that other aspects aside from cost are better considered in the evaluation of DB best-value projects.

The lack of influence of non-cost factors might be due to the inability of evaluation criteria to elicit meaningful and consistent information to evaluate and compare the proposals. Design-builders cannot offer the best-value in their proposals if evaluation criteria do not show precisely what constitutes best-value and how best-value is scored. In this context, best-value refers to the greatest benefit that a proposer might provide for a particular project (AGC of America and NASFA 2008). In US design-build highway projects procured using best-value, the benefits that the proposer might provide usually relates to schedule, technical merit, management options, and past performance (Molenaar and Tran 2015).

Challenges in eliciting meaningful information as well as lack of clarity in what constitutes best-value might be related to evaluation criteria not being written consistently with decision-making theory, which provides specific formulations (i.e., direct scoring or constructed scale) and characteristics (i.e., comprehensiveness, directness, unambiguity, and understandability).

Previous research has primarily examined evaluation criteria under two approaches. First, several studies addressed the issue of what criteria should be considered in the best-value evaluation (De_Araújo et al. 2017; Gransberg et al. 1986; Montalbán-Domingo et al. 2019; Scott et al. 2006; Shalwani et al. 2019; Xia et al. 2011a; b). Second, few studies and industry associations have focused on how criteria should be articulated and written (AGC of America & NASFA 2008; Gransberg et al. 2006; Molenaar and Tran 2015).

As part of the first group, Gransberg et al. (1986) listed four essential types of criteria that should be considered in any design-build project. These were management, schedule, technical, and cost. Scott et al. (2006) categorized best-value criteria in management, schedule, cost, and design alternate. More recently, De Araujo et al. (2017) found through a systematic literature review that for highway projects, the most used criteria were cost, time, quality, staff features, and financial. Xia et al. (2011a) identified price, experience, management, and qualifications as being the criteria most widely included in RFP of design-build highway projects. A recent study by Shalwani et al. (2019) analyzed 362 bidding results to determine what evaluation criteria had the most significant differentiation in scores for competing bidders. The focus of these studies remains on the identification of evaluation criteria used in design-build best-value, but not on their formulation. This research fills this gap by providing a structured approach to authoring best-value evaluation criteria.

Concerning how agencies define evaluation criteria, Gransberg et al. (2006) highlighted the relevance of "well-written criteria"—that must be clear, unambiguous, and definitive—in the development of design-build proposals. Similarly, the guide developed by AGC of America and NASFA (2008) on *Best practices for best value selections* recommends that evaluation criteria "should yield in-depth information from the contractors on their specific approach to delivering the expected service." This is, evaluation criteria must be able to communicate what constitute best-value. Otherwise, proposers approach the procurement with a low-bid mentality (Tran et al. 2018). Molenaar and Tran (2015) studied best-value practices across the US and suggested that evaluation criteria need to be clear, easy to understand, and project-specific; criteria should define how each agency will score them. These recommendations are aligned with AASHTO (2008) and DBIA (2019) guidelines, which highlight that evaluation criteria must be measurable. They also suggest that RFPs need to clearly articulate the basis for evaluating best-value design-build proposals. Previous studies analyzing the definition of evaluation criteria have derived general recommendations on how criteria should be written (i.e., should be clear, well written, etc.) However, these studies neither analyze the specific formulation of a large set of evaluation criteria nor use a supporting theory to establish their recommendations.

The research addresses this gap by providing a structured approach for writing best-value evaluation criteria on design-build highway projects. This approach is grounded in established decision analysis theory of evaluation criteria. And, it is applied to the practice of best-value procurement in design-build highway projects in the US.

While this paper will be of interest to researchers and practitioners, the primary contribution is to improve agency practice in the authoring of best-value evaluation criteria.

BEST-VALUE PROCUREMENT CRITERIA THROUGH THE LENS OF DECISION ANALYSIS THEORY

Best-value evaluation criteria are used by highway agencies to assess the requirements (i.e., information) presented by the design-builders and measure the extent to which each design-builder meets the agency's objectives. The ultimate goal of any decision-making process is to select the alternative that best meets the decision-makers' objectives. The ultimate goal of best-value procurement is to select the design-builder that best meets the agencies' objectives by providing the overall best-value to the projects. With objectives, requirements, and evaluation criteria tightly related, the formulation of best-value evaluation criteria should build upon the established objectives and requirements.

Theoretically, decision analysis considers the decomposition of decision problems into the choices, information, and preferences of the decision-maker (Howard 1980). In other words, decision analysis prescribes how a decision-maker should systematically think about structuring the decision problem (Keeney and Raiffa 1976).

Best-value procurement constitutes a decision problem with part of it being the articulation of evaluation criteria. Thus, rigorous and validated decision analysis approaches can help to conduct a systematic analysis of highway agencies′ current practice on authoring evaluation criteria. To this end, the following sections elaborate on the characteristics that requirements and evaluation criteria should systematically hold to produce consistent selections.

Requirements

In best-value procurement, the RFP's state the requirements regarding the format, length, and level of detail of information that should be included in the proposals (AASHTO 2018). As in any rigorous decision-making problem, all relevant information in the context of the selection should be included, and this information should be easy to analyze (Clemen and Reilly 2000). High-importance requirements should be measurable and unambiguous (PMI

2016). A measurable requirement implies the use of evaluation criteria to assess the degree of accomplishment of objectives. In this regard, it is essential to analyze the type and the amount of information that is going to be required. An unambiguous requirement, on the other hand, has a single meaning and is interpreted the same way by any audience.

The requirements should be established by considering both the objectives and the evaluation criteria. Objectives will determine the content of the requirements, whereas the evaluation criteria will determine how the content should be required to be appropriately measured.

Evaluation criteria

Evaluation criteria constitute a set of measures that describe the contribution of each design-builder to accomplishing the agency's objectives (Keeney and Gregory 2005). As a metric, they possess four primary features: 1) having a need or a purpose: 2) providing useful information; 3) focusing toward a target and; 4) being able to be measured with reasonable accuracy (Kerzner 2017).

Formulation

Evaluation criteria can be classified based on their relationship with the objective (Keeney and Gregory 2005). A natural evaluation criterion can directly measure the objective of concern. For example, if the objective is to minimize cost, the evaluation criteria "cost in dollars" is a direct measure of the objective (Keeney and Gregory 2005). In the case of best-value evaluation criteria, this research considers that an evaluation criterion is natural if it can be measured using a direct scoring. In case that a natural evaluation criterion does not exist to measure the objective directly, it is possible to build a constructed one by defining a scale where the different levels of accomplishment of the objective can be measured (Clemen and Reilly 2000; Keeney and Gregory 2005; Keeney and Raiffa 1976). This might be the case of best-value evaluation criteria that are defined by distinct levels—e.g., excellent, good, and moderate—to differentiate the degree of accomplishment of the objectives based on the requirements.

Characteristics

Seminal work on decision analysis theory highlights the importance of evaluation criteria being comprehensive, direct, unambiguous, and understandable (Belton and Stewart 2002; Keeney and Gregory 2005; Keeney and Raiffa 1976). An evaluation criterion is comprehensive if the decision-maker has a clear understanding of the "extent that the associated objective is achieved" (Keeney and Raiffa 1976). An evaluation criterion is direct when its levels directly describe the consequences for the objectives of interest (Keeney and Gregory 2005). Evaluation criteria guidelines in other fields (i.e., technology services, health or information, and image management) refer to being comprehensive as "being able to separate best, average and weaker proposals." These guidelines also point to direct evaluation criteria as: "all key elements of the project requirements must relate to the requirement definition and be covered by evaluation criteria" (NCOITS 2008); "The evaluation criteria must address all key elements of the requirements" (Porter-Roth 2007); and "All key elements of the project requirements must be covered by evaluation criteria." (UTHealth 2020)

Highly related to being direct, effective evaluation criteria must also be unambiguous. When an evaluation criterion is unambiguous, it is possible to precisely describe the result of the assessment using the evaluation criteria (Keeney and Gregory 2005). Evaluation criteria guidelines in fields such as technology services and health emphasize this feature, recommending evaluation criteria to be clear, objective, and not subject to multiple interpretations." (NCOITS 2008; UTHealth 2020). Some highway agencies concern about the subjective nature of best-value evaluation (Chini et al. 2018). Unambiguous evaluation criteria will led to minimize this concern.

Finally, evaluation criteria must be understandable to ensure that all the stakeholders involved in the decision have a shared understanding of the concepts that will be used in the selection process (Belton and Stewart 2002; Keeney and Gregory 2005). Understandable evaluation criteria are fundamental for clear communication of the pros and cons of the different alternatives (Keeney and Gregory 2005).

Figure 1 summarizes the concepts shown in this section and will serve as a framework for analyzing and proposing a consistent approach for writing best-value evaluation criteria.

< FIGURE 1>

Decision analysis theory guides how to formulate evaluation criteria and the desirable characteristics that criteria should have to reach a consistent best-value selection. This research uses decision analysis as a framework to evaluate and enhance highway agencies' practice in writing best-value evaluation criteria. This approach fills the gap of current research in this regard. Previous studies neither analyzed the specific writing practices of a large set of evaluation criteria nor used a supporting theory to establish their recommendations.

RESEARCH APPROACH

This research applied a structured approach to create best-value evaluation criteria that is founded in the development of a requisite decision model. A requisite model is a simplified representation of reality and can be defined as "a model whose form and content are sufficient to solve a particular problem" (Phillips 1984). Requisite models differ from descriptive models because the goal of requisite models is to serve as a guide to action, whereas descriptive models roughly tell what people actually do. Requisite models are generated by the interaction between specialists and problem owners. Specialists contribute to the form of the model, whereas problem owners provide content (Phillips 1984). In this research, the requisite model and resulting structured approach are grounded in established decision analysis theory (which defines the form of the model) and the content is provided by current practice on evaluation criteria (extracted from 540 evaluation criteria in published best-value RFPs). The interactions between these two components (i.e., form and content) consist of an iterative process to ensure the proposed approach reflects the needs of highway agencies while being grounded in decision analysis theory. The concept of a requisite model being applied to the structure of a decision problem implies that the model can be " a structural representations simple enough to capture the essence of a decision problem, and no more complicated than necessary to obtain sound insights" (Winterfeldt and Fasolo 2009).

Requisite models have served to address different purposes in previous construction management research. For example, in enhancing knowledge construction processes with multicriteria decision analysis (Vieira et al. 2020); modeling trade-off between overlapping and rework of design activities (Dehghan and Ruwnapura 2014), selecting the drivers in project delivery method's selection (Touran et al. 2011); or facilitating bid evaluation in public calls for tenders (Bana et al. 2002)

In this research, the goal of the requisite structured approach is to guide highway agencies in authoring best-value evaluation criteria. It demonstrates how these evaluation criteria can be explicit about what constitutes best-value and how best-value is scored. The ultimate goal is to show the ways that evaluation criteria can be change so that they are more meaningful, or influential, in the best-value evaluation outcome.

To develop the content of the requisite model, the researchers conducted a deductive content analysis on request for proposals (RFPs) of design-build highway projects. These projects were procured using a best-value method across the US. Previous research also used content analysis to analyze the frequency of use of different categories of evaluation criteria (Xia et al. 2011b; a). Through deductive content analysis, researchers structure the analysis based on previous knowledge or theory (Elo and Kyngäs 2008). In this case, the theory used was the decision analysis theory, given that this theory helps to structure the definition of all the elements that influence the decision (Howard 1966). The types of evaluation criteria formulation defined by Clement and Reilly (2000) and Keeney and Gregory (2005) were the initial categories for the analysis of the evaluation criteria included in the RFPs. According to Catanzaro (1988), deductive content analysis is regularly used when researchers aim to evaluate existing data in new contexts. This research uses this approach because its purpose is to assess and improve current best-value evaluation criteria formulation under the lens of decision analysis theory.

The researchers used theoretical saturation to determine the number of projects and evaluation criteria to be analyzed. According to Saunders et al. (2018), saturation means that without additional data, the researcher can develop theories of a category and that seeing similar instances over and over again, the researcher "become empirically confident that a category is saturated." The initial categories in this research for analyzing the evaluation criteria were "direct scoring," "constructed scale," and "neither of them." For each project, the researchers considered the evaluation criteria formulation from the RFPs and analyzed whether their formulation corresponded with one of the three defined groups. In case the excerpted evaluation criteria did not match with any of the groups, a new category would be created. The researchers followed this process from state to state and project to project until categories were found stable and supported by sufficient data.

For the analysis, the researchers ranked the states based on their expenditure on highway projects (FMI 2018; Tax Policy Center 2017). Fourteen states that did not have specific authorization or had certain limitations to using design-build, according to DBIA (2017), were not considered. Preliminary results derived from the three states having the highest expenditure in highway projects showed that current evaluation criteria did not follow the theoretical "constructed scale" formulation. Rather, RFPs included generic constructed scales, where the levels, instead of relating specifically to the requirements of each evaluation criteria, were common for all of them. Thus, researchers added the category "generic constructed scale." In addition, those evaluation criteria that were neither direct scoring nor constructed scale had a common characteristic. All of them were generic expressions that did not establish any relationship between the evaluation criteria's scores and how these scores would be assigned depending upon the requirements' comprehensiveness. Thus, the researchers changed the name of this group from "neither of them" to "generic expression."

The analysis was conducted on a total of 21 states, accounting for a total of 98 projects and 540 evaluation criteria. All the evaluation criteria corresponded to one of the three categories determined in the first iteration: “direct scoring," “generic constructed scale,” and “generic expression.” Thus, the researchers considered the sample representative for the categories to analyze.

Table 1 lists the states considered and the number of projects per state. Data were collected from procurement documentation available in the Department of Transportation (DOT) websites. The order shown is the one followed by the researchers in the analysis.

< TABLE 1>

Overall, the analysis followed a three-step approach. First, the evaluation criteria were classified in one of the categories defined: “direct scoring," “generic constructed scale,” or “generic expression.” Second, the evaluation criteria were analyzed to determine to what extent they were comprehensive, direct, unambiguous, and understandable. Third, based on the evaluation criteria groups and characteristics, the research suggests a structured approach to write best-value evaluation criteria. The following section “analysis of current practice” describes the first and second steps of the methodology, whereas the section “structured approach for writing consistent criteria” covers the third one.

ANALYSIS OF CURRENT PRACTICE

The formulation of the evaluation criteria found in the current state of the practice varies between three different categories. First, several RFPs define evaluation criteria by setting a list of requirements and an associated score, but without detailing how that score relates to the levels of accomplishment of the requirements. Strictly, in these cases, the evaluation criteria are not defined—the research classified these evaluation criteria as “generic expression.” Second, in the cases where the evaluation criteria were identified using a constructed-scale approach, this scale was vaguely defined. In these two initial groups, evaluation criteria were not comprehensive, direct, unambiguous, nor understandable. Finally, few cases follow a direct scoring approach, which generally was comprehensive and direct and just needed some improvements to become completely unambiguous and understandable.

The following sections describe each of the three evaluation criteria formulations found in current practice. Each section includes examples and discussion about their desirable characteristics. These examples were chosen because they are representative of the issues that were discovered in the analysis.

Generic expression

Forty-three percent (43 %) of the 540 evaluation criteria analyzed were formulated using a general expression. Among them, it was also possible to differentiate distinct formulations. The most generic one refers to evaluation criteria that are defined as requirements with an associated score. This type of evaluation criterion uses verbs such as *describe, list*, *provide*, *submit*, etc. Each requirement then has a score associated with it. But details are not included about how the score should be assigned depending on the level of accomplishment of the requirements.

Other generic formulations use expressions such as *this evaluation criterion will measure, the degree to which*, *the effectiveness*, *the extent to which*, or *credits will be given*. In other cases, the evaluation will be based on the likelihood and degree to which the design-builder's commitments will *achieve*, *minimize*, *demonstrate ability*, or *demonstrate efficiency*. These expressions are accompanied by the related score, but there is no guidance on how the score reflects different levels of accomplishment.

The following representative example of this type of generic evaluation criteria expression will support the analysis of the four evaluation criteria characteristics: comprehensiveness, direct, ambiguity, understandability.

- Evaluation criterion: Management/Administration
- Requirement: Preliminary Project Management Plan
- Formulation: *The department will use the following evaluation criterion to score the management portion of the technical proposal: The degree to which the Preliminary Project Management Plan (PPMP) demonstrates an efficient approach to the management of traffic during the Construction Period.*

(Adapted from Caltrans RFP I-15/I-125)

This evaluation criterion is neither comprehensive nor direct because the expression "the degree to which" does not specify the levels of assessment. Given that these levels are not clearly stated, it is not possible to make a direct relationship between score/level and the requirement's characteristics. Further, the evaluation criterion in the example is ambiguous because the expression "demonstrates an efficient approach" is not specific and does not define what an "efficient approach" means. Finally, this criterion is not understandable because the expression "the degree to which the PPMP demonstrates an efficient approach" does not establish a direct relationship between the different levels of PPMP efficiency and the related scores. Therefore, there is not a unique meaning associated with each different score.

This example shows a case where one requirement is evaluated by one evaluation criterion. However, the common practice is to have several requirements that are evaluated by the same evaluation criterion or an evaluation criterion that assesses different aspects of the same requirement. In these cases, not including any prioritization of the distinct evaluation targets makes the evaluation criteria even less comprehensive, direct, unambiguous, and understandable.

Generic constructed scale

Fifty-three percent (53 %) of the 540 analyzed evaluation criteria were formulated with a generic constructed scale. In these cases, the levels of accomplishment are defined in general terms and do not consider the specific requirements asked for each project. A generic scale makes it challenging to differentiate the levels of accomplishment based on the requirements. Also, several requirements and a generically constructed-scale hinder the establishment of a direct relationship between these requirements and the levels of the scale. Among the evaluation criteria using a generic scale, it was found three variants that are explained through the following examples.

The first type of evaluation criterion includes a scale with few levels and a very generic definition of each level. The following is a representative example:

- Evaluation criterion: Collaboration
- Requirements: (1) Provide a narrative describing tangible examples of effective issue resolution (…); (2) Describe examples where the design-builder has approached project challenges with a collaborative attitude. (3) Describe how the design-builder built trust with the owner (…).
- Formulation:
    - *Good Range: approach that generally meets the RFP requirements.*
    - *Very Good or Excellent range: proposal including specific approaches and/or specific commitments that are considered to exceed the RFP requirements, such as providing advantages, benefits, or added value to the project; reducing and/or avoiding risks; minimizing cost and/or schedule impacts; and resolving issues in the best interest of the project. Also, proposal might receive this rating if they cite recent examples of successful partnering and references confirm certain aspects of the evaluation.*

(Adapted from WSDOT Interchange Direct Connector)

This evaluation criterion attempts to define a scale that differentiates the levels of accomplishment. However, three main aspects prevent it from being completely comprehensive. First, the low number of levels (good/very good or excellent). Second, the ambiguity in the definition of these levels using expressions such as “very good or excellent" “generally meets/considered to exceed the RFP requirements.” Third, the inclusion of several aspects within the same level linked by “and/or.” The evaluation criterion is not direct because there could be several proposals included in the “very good” range with different levels of content. Further, this example is ambiguous because the expressions used, such as *generally meet the requirements* or *are considered to exceed the requirements* do not refer directly to the specific requirements. It is neither understandable because if a decision-maker says this proposal is excellent, it is not possible to know precisely what aspects make it different from others.

The second type of evaluation criterion in this group represents the most common formulation, characterized by evaluation criteria, including a constructed scale defined in broad terms. The following is a representative example:

- Evaluation criterion: Construction Staging and Traffic Management Plan
- Requirement: Describe the construction staging and traffic control and sequencing proposed to accommodate and minimize impacts to traffic (…)
- Formulation:
    - *The degree to which Design-Builder’s preliminary Construction Staging and Traffic Management Plan utilizes a safe, effective strategy to minimize the Maintenance of Traffic (MOT) impacts to corridor motorists and reduce any lane or shoulder closures required.*
    - *The degree to which the Design-Builder utilizes innovative technologies to minimize impacts to the traveling public.*

    Unless otherwise, the proposals will be scored using qualitative/descriptive rating methods, as is summarized in Table 2”.

< TABLE 2>

(Adapted from GDOT I-85 widening)

In this case, the evaluation criterion is comprehensive because it is possible to distinguish different levels of accomplishment. However, it is not direct. Although the evaluation criterion defines levels according to the degree in which the proposals meet the requirements, it is not clear, for example, how to *exceed in a significant manner* differs from only *exceeds*. The levels of the generic constructed scale do not relate specifically to the requirements and the evaluation criteria. Thus, it is difficult to apply this scale to assess the requirements according to the evaluation criteria. For example, the assessment of the requirement "description of the construction staging, traffic control, and sequencing to accommodate and minimize impacts to traffic" under the evaluation criterion "the degree to which the design-builder utilizes innovative technologies to minimize impacts to the traveling public" cannot be done directly by using the levels defined in the scale.

Further, the generic definition of the scale's levels might lead to different interpretations depending upon the person who is conducting the assessment. The scale is defined in general terms and makes the evaluation criterion ambiguous. Finally, if a decision-maker says that this proposal is excellent in the evaluation criterion *construction staging and traffic management plan*, it is not possible to know precisely what aspects make it different from others in regards, for example, innovative technologies application. For this reason, this criterion is not understandable.

The third type of evaluation criterion considered in this section is a singular case. It is worth mentioning because it includes a constructed scale but also a prioritization of requirements. Further, it provides clarification about the meaning of some of the language contained in the definition of the levels.

- Evaluation criterion: Safety and Mobility
- Requirements:
    - (1) Provide a narrative that describes your project and discuss how your project maximizes the number of continuous four-lane dualized roadway (…) Provide a discussion of the project elements (…) Include a discussion of any approved Alternative Technical Concept (ATC) (…). This requirement is critical.
    - (2) Describe how the project will improve network traffic operations and reduce crashes. Include a discussion of any qualitative and/or quantitative analyses. This requirement is significant.
    - (3) Identify any conditions in your project that do not meet the 10 AASHTO Controlling Criteria and describe how your project will mitigate these conditions. This requirement is important.

Language clarification

- “Critical” requirements are approximately three times the relative importance of “important” ones.
- “Significant” requirements are approximately two times the relative importance of “important” ones.

- Formulation:
- Table **3** shows how this evaluation criterion is formulated.

< TABLE 3>

Language clarifications:

- The term “weakness” means any flaw in the proposal that increases the risk of unsuccessful contract performance.
- A “significant weakness” in the proposal is a flaw that appreciably increases the risk of unsuccessful contract performance.
- The term “deficiency” means a material failure of a proposal to meet an RFP requirement or a combination of significant weaknesses in a proposal that increases the risk of unsuccessful contract performance to an unacceptable level.

(Adapted from MDOT RFP MD 32 to I70)

This evaluation criterion is comprehensive because the constructed scale enables decision-makers to distinguish the different levels of accomplishment of the requirements. The constructed scale in the example is not explicitly built for the evaluation criterion and its requirements—i.e., the same constructed scale is used for all the evaluation criteria in the RFP—therefore, it is not completely direct. However, this example is more direct than the previous one in Table 2. This case’s levels focus on the assessment of the requirements on four specific concepts: understanding, quality, risks, and strengths/weaknesses. In contrast, the levels in the case in Table 2 assess in general terms—e.g., exceed, conformably meet—how well the requirements were met.

This evaluation criterion is ambiguous because it is unlikely that different people reach the same interpretations about what is measured. Although the scale is focused on evaluating understanding, quality, risks and strength/weaknesses, it would be necessary to specify, for example, what differentiates “complete understanding” from “strong understanding” or what defines “highly skilled team," “experienced team” or “qualified team” in regards to quality accomplishment. The evaluation criterion is also not understandable. If a decision-maker says this proposal is exceptional in the evaluation criterion *safety and mobility*, it is not possible to know precisely what aspects make it different from others in regards to the improvement of traffic operations and reduction of crashes.

Overall, this group of representative examples shows how evaluation criteria can measure the requirements with more detail than the ones included in the “generic expression” category. The “generic constructed scale” category provides a structure of levels where the different degree of requirement’s accomplishment can be distinguished. However, this scale fails to be specific and to establish a direct relationship between the grade of requirements’ contents and the scale levels. Thus, these types of evaluation criteria are neither direct, unambiguous, nor understandable.

Direct scoring

Four percent (4 %) of the 540 analyzed cases use specific rules for assigning direct scoring to specific requirements. This group considers fewer requirements than the other groups, and the evaluation criteria include prioritization of them and direct assignment of the points. Two representative examples are shown below for illustrative purposes.

Example 1

- Evaluation criterion: Traffic Performance
- Requirement: “Synchro” models based on the design.
- Formulation: the maximum quality evaluation points are distributed as Table 4 shows:

< TABLE 4>

“*The design-builder with the fastest total sum of peak hour travel times for a particular roadway segment(s) will receive the maximum points for that segment. Remaining design-builder’s time to be pro-rated against the fastest time. For example, For the Road Segment: NB Route 32 Corridor, the calculations for assigning the scores are included in Table 5*”.

< TABLE 5>

*“Design-builder A has the fastest total sum of peak hour travel time at 8.98 minutes for the NB Route 32 corridor, thereby receiving 5 points. Design-builder B, with a total sum of peak hour travel time of 9.31 minutes, would receive 8.98/9.31*5=4.82 points”.*

(Adapted from NY DOT RFP Route 17)

Example 2

- Evaluation criterion (defined by its associated objective): Minimize impacts and inconvenience to the community, motorists, businesses, downtown, and the public during construction.
- Requirements: The previous objective is measured by using the following information (1) Project Completion Deadline and (2) Maintenance of Traffic.
- Formulation: of the sub-evaluation criteria: project completion deadline.

  (1) Project completion deadlines. The equation that will be used is:

  "*Milestone Duration Points=MxPts * (CDR/CDRmx)*

  MxPts =Maximum allowed points milestone duration

  CDR =Design-builder's Calendar Day Reduction

  =Maximum allowed Calendar Days-committed Calendar Day in Form P

  CDRmx =Calendar Day Reduction of the Design-builder with the shortest schedule for the Milestone Duration."

(Adapted from CODOT RFP Cimarron)

In these cases, the evaluation criteria are comprehensive and direct. It is possible to differentiate the levels of accomplishment of each alternative in the evaluation of the requirements. Further, it is possible to relate scores with the requirements features; for example, in the first case, acknowledging the score, it is possible to know the associated travel time. Also, generally, these evaluation criteria are unambiguous and understandable because different people can reach the same interpretation of what is measured. The meaning of the score is clear.

Summary of current practice

The analysis of current practice revealed that in 43% of the cases, evaluation criteria are set using a general expression, which is neither direct scoring nor constructed. In 53% cases, evaluation criteria are defined using a generic constructed scale, which does not allow measuring precisely the related requirements. Finally, 4% of cases, evaluation criteria are specific and use direct scoring for assessing the requirements.

This research found that a high percentage of the evaluation criteria included in the best-value RFPs in this study do not follow a structured formulation such as direct scoring or constructed scale. This is a key limitation because, as the analysis shows, the more generically defined an evaluation criterion is, the less capable of being comprehensive, direct, unambiguous, and understandable (Figure 2)

< FIGURE 2>

STRUCTURED APPROACH FOR WRITING CONSISTENT CRITERIA

The path towards improving current practice on formulating best-value evaluation criteria moves from using general expressions to defining specific direct score or constructed scale evaluation criteria. The authors suggest an approach to transition from current practices to one that is based on proven decision-making principles and analysis of existing RFPs. It comprises ten questions that relate to the selection objectives, the requirements, and the evaluation criteria characteristics. This approach guides the creation of consistent evaluation criteria that can be adapted to the various state highway agency evaluation criteria. The structured approach proposes specific questions depending on the current evaluation criteria group—i.e., generic expression, generic constructed scale, or direct scoring (Figure 3). This structure is the result of the iterative process followed to ensure the proposed structured approach reflects the interactions between decision theory analysis and current practice on evaluation criteria. Depending on where the evaluation criterion lies in the evaluation criteria spectrum (i.e., generic expression, generic constructed scale, or direct scoring) the proposed approach suggests a different set of questions aimed at enhancing its formulation.

<FIGURE 3>

Each of the formulations found in the current practice represents different levels of specificity in writing evaluation criteria. As this research shows, the formulation of the evaluation criteria is tightly related to the definition of objectives and requirements. For this reason, the set of questions proposed cover objectives (questions 1 and 2), requirements (questions from 3 to 5), and evaluation criteria (questions from 6 to 10). Specifically, questions 6-10 aim to check if the evaluation criteria are comprehensive, direct, unambiguous, and understandable, respectively.

The generic expression is the least specific evaluation criterion formulation. It does not follow any of the proposed formulations of constructed scale or direct scoring. Thus, this research suggests highway agencies having generic expression formulations to start in question 1.

The generic constructed-scale falls at a medium level of specificity. It follows the constructed-scale formulation but fails in not being specific for each objective and associated requirements. Usually, RFPs include one generic constructed-scale that serves to evaluate all the different requirements. This type of criteria formulation does not establish a specific link between the description of their levels and the content of the requirements; as a consequence, these criteria are not direct. For this reason, this research proposes highway agencies having generic constructed-scale formulations to start in question 6.

Finally, direct scoring constitutes the most specific evaluation criteria formulation. It follows the direct scoring formulation but might have problems with being unambiguous and understandable if an abstract language is used in the rules of score assignment. To address this issue, this research proposes highway agencies having direct scoring formulations to start in question 9.

To illustrate the application of the structured approach in the three types of evaluation criteria formulation, the following section provides actual examples and their re-formulation based on the suggested approach.

Generic expression example

Table 6 includes an example of generic expression criteria related to safety. This example will guide the application of the suggested structured approach in this group of evaluation criteria.

<TABLE 6>

Question 1 points out the identification of the objectives. In the example provided, the decision-makers aim to measure how the design-builders provide safety conditions to the workers, DOT, and other people in the project's area. They measure it by looking at the final and specific design-builder's commitments. Question 2 refers to what information is required for the design-builders to measure the objectives. In this case, four pieces of information are required: an approach to mitigating safety, approach to proactively enhancing safety practices, final commitments, and specific commitments. However, only the commitments will be evaluated regarding safe working conditions. Decision-makers should examine if this is the appropriate information to require for measuring and differentiate design-builders in regards to providing safe conditions.

Questions 3 and 4 refer to the requirements' suitability to be assessed by the evaluation criteria. This example considers two requirements that are not going to be explicitly evaluated (approach to mitigating safety and approach to proactively enhancing safety practices), which could be misleading. The other two are commitments, which are differentiated in two types: final and specific. The evaluation criterion assesses both of them under the view of providing safe conditions. In this case, the requirements are general; thus, an accurate assessment is difficult. Question

5 helps to think about how the requirements will be assessed. In this case, it is necessary to describe more precisely the requirements, specifically, the definition of commitments, so that it could be possible to asses them using a direct score or a constructed scale. Overall, the analysis of the requirements should be made by examining both objectives and evaluation criteria. Objectives determine the content of the requirements while the evaluation criteria define how the content should be required to be correctly measured.

Question 6 asks if it is possible to differentiate the levels of accomplishment of each alternative in the evaluation of the requirements. In the example, the answer is no. The evaluation criterion only indicates that “the approach will be evaluated based on the likelihood and degree to which the design-builder’s commitments will achieve” safe working conditions for the stakeholders involved in the construction. It is necessary to think about how “safe working conditions” will be measured in the “design-builder's commitments." Further, it is key to define how the rules for assigning different scores/levels relate to the "degree" of safety conditions that each design-builder includes in his/her commitments. Formulating the evaluation criteria using a constructed scale or direct scoring makes the evaluation criteria to be comprehensive. The first step is to create the levels of requirements' accomplishment and the rules of scores' assignment. To this end, question 7 might help. This question refers to what constitutes the best/average/worst requirements that a design-builder could submit.

By asking the previous questions, the evaluation criterion is re-formulated, as Table 7 shows. The requirements are also re-formulated based on two criteria: (1) impact on project cost and schedule and (2) impact on injury rates, near misses, and event with significant injuries. Depending on the impact's’ value on these two criteria, different levels of requirement's accomplishment were defined. These levels determine what best, average, or worst is, making a comparative evaluation of the impacts' value included in each proposal. The evaluation criterion states clearly that the proposal's "best-value" will be measured based on their ability to reducing injury rates, near misses, and events with significant injuries. If two proposals are similar in this regard, the best one will be that proposal whose commitments generate a minimum impact on the project cost and schedule.

< TABLE 7>

This evaluation criterion is comprehensive because it is possible to distinguish the design-builder's degree of accomplishment. The next step is to check the other three evaluation criteria characteristics (direct, unambiguous, and understandable) by asking questions 8, 9, and 10. An evaluation criterion is direct if given one design-builder's score or level, it is possible to know to what extent a proposal reaches the requirements. In the re-formulated example, this condition is achieved. If a decision-maker says that the safety portion of the proposal is good, it is possible to know the characteristics of the requirements directly: 10 commitments that demonstrate an average reduction (among all the proposals) in injury rates, near misses, an events with significant injuries.

An evaluation criterion is unambiguous if different evaluators reach the same interpretation/measurement (question 9). In this case, this is also true. Everyone can achieve the same understanding of what is an excellent, good, fair, or poor safety proposal. This is because the constructed scale is built based on specific information linked to the requirements.

Finally, an evaluation criterion is understandable if the language used to assign the scores/levels is precise (question 10). This re-formulated evaluation criterion is understandable because the levels are explicitly defined in terms that unambiguously relate to the requirements.

This section constitutes a representative example of how the guidance of a structured approach might transform generic expressions into consistent evaluation criteria.

Generic constructed scale example

Table 8 includes a representative example of generic constructed-scale criteria related to safety. This example will guide the application of the suggested structured approach in this group of evaluation criteria.

< TABLE 8>

This type of evaluation criteria already has a constructed scale where it is possible to distinguish the different levels of accomplishment of each alternative regarding these levels. However, the connection between the description of the levels and the definition of the requirements is not direct. Questions 6-7 might help to analyze the content of the requirements and their relationship with the levels of the evaluation criterion. The requirements, in this case, are two: (1) safety considerations and (2) firm's overall approach to safety. On the other hand, the description of the evaluation criteria levels mentions four concepts: (1) Requirements: Significantly exceed/exceed/meet; (2) Quality: Provide consistently outstanding level/better than acceptable/acceptable of quality; (3) Strengths: Significant/some and; (4) Weaknesses: No/no significant/minor or moderate/significant weaknesses. These descriptions make it challenging to know the level of accomplishment of the requirements. For example, what does it mean to exceed the safety considerations requirement? What does it mean to provide a better than acceptable level of quality? It is not clear. The levels of accomplishment of the requirements are not clearly defined.

In the example, given that the levels of accomplishment are not clearly defined, if one decision-maker says, "This proposal's safety considerations have a score of 95", it is not possible to know the characteristics of the requirements that the design-builder provided (question 8). It is not possible to understand what the proposal included to exceed the requirements significantly and to provide a consistently outstanding level of quality. Further, in this case, there are two requirements associated with the evaluation criteria, but there is not any information about if one of them is more important than the other. There is no prioritization. For these reasons, this evaluation criteria is not direct. To make this evaluation criterion direct, it is necessary to make the requirements more specific, prioritize them, and establish a direct relationship between these requirements and the description of the evaluation criterion's levels.

By asking questions 6, 7, and 8, the evaluation criterion can be re-formulated, as shown in Table 9. In this case, instead of prioritizing the requirements, the firms' "overall approach to safety" was considered to be a pass/fail evaluation criteria, meaning it is considered not possible to score. The requirement "safety considerations" was formulated in terms of identifying risks and including specific information about them. The evaluation criteria described in each of the levels relate directly to the requirements. The criteria prioritize the relevance of the risks, their influence's justification and the cost-effectiveness ratio of the proposed mitigation measures.

< TABLE 9>

This evaluation criterion is comprehensive and direct. The next step is to check the other two evaluation criteria characteristics (unambiguous and understandable) by asking questions 9 and 10. An evaluation criterion is unambiguous if different evaluators reach the same interpretation/measurement (question 9). In this case, this is almost true. Everyone can achieve the same understanding of what is in the score range of 90-100 or 70-79. This is because these ranges are built based on specific information linked to the requirements. However, the decision-makers should create another rule for determining how they would assign the scores within each level (e.g., what is the criteria for assigning 97 instead of 94 in the 90-100 range). The process would be similar to the one already showed but focusing on more detail of evaluation.

Finally, an evaluation criterion is understandable if it uses precise language in the definition of point's assignment (question 10). This re-formulated evaluation criterion is understandable because the levels are explicitly defined in terms that unambiguously relate to the requirements.

Overall, this section constitutes a representative example of how following a structured approach might transform generic constructed scales into consistent evaluation criteria.

Direct Scoring example

Table 10 shows an example of direct scoring criteria related to safety, capacity, and operation. This example will guide the application of the suggested structured approach in this group of evaluation criteria.

< TABLE 10>

This evaluation criterion is comprehensive and direct. The relationship between the requirements and the points that are assigned is clear. In this case, the key questions are the ones related to checking the unambiguity and understandability of the evaluation criterion (questions 9 and 10).

Overall, this example is unambiguous. However, it fails in a small aspect of its definition, which is the assignation of points to ARE 4. It reads: "ARE 4 will be variably scored up to a maximum of 7 points based on its effectiveness at improving additional capacity and operations improvement (…)". When evaluation criteria introduce abstract language such as "effectiveness," they are also introduction ambiguity. How is "effectiveness" measured? In regards to this evaluation criterion, different evaluators can reach different interpretations of what is measured.

In order to make this part unambiguous, the paragraph associated with the evaluation of ARE 4 could be reformulated as follows (Table 11):

<TABLE 11>

The re-formulation proposes that ARE 4 will be variable score up to a maximum of 7 points based on two parameters: (1) reduction of waiting time in peak hour (RWT) and (2) average time savings per day (TTS). Both of them using the "X" traffic simulation model and comparing the scenario with and without ARE 4. With this change, the whole evaluation criterion becomes understandable because it is clear what any score from 0 to 25 means.

VALIDATION OF THE PROPOSED APPROACH

The validation of descriptive models is a general practice among the research community and relies on the correlation between actual data and predicted outcomes from the descriptive models. In descriptive modeling, high correlations are considered indicative of model validity. This is, however, not the case for requisite models. According to Philips (1984), validating a requisite model requires the development of a requisite evaluation model. The validation of decision models is indeed an unresolved problem within the scientific community because decisions do not model any physical reality (Collier and Lambert 2019). According to Gass (1983), a decision model could at best be partially validated, because researchers will never have full data of the alternatives not selected.

The validation of decision models is generally related to the model's utility to provide insights (Gass 1983). Greenberg (1988) described this validation as the extent to which the model can generally lead to good decisions and keep away bad decisions. According to Howard (1966), the only way we can assess the quality of a decision is by whether it is consistent with the choices, information, and preferences of the decision-maker. The quality of the decision is, therefore, not the same as the quality of the outcome. Making a good decision means to do the best it is possible to increase the chance of a good outcome (Howard 1983).

The validation of the proposed approach also relies in the integrity of the method followed to create the requisite model, whose form is founded in decision analysis theory and whose content is provided by actual RFPs. According to Lucko and Rojas (2009), one of the most relevant ways that researchers can show the integrity of a research methodology is documenting the entire approach in detail with an open and self-critical mind. In this line, Creswell (2009) suggests qualitative validation as one strategy. This involves the use of a comprehensive description of the procedure followed by the researchers to convey the findings. The research approach section includes a detailed description of the process followed. This description explains: (1) why decision analysis theory was used (to provide an evaluation criterion proven structure); (2) how current practice was analyzed (content analysis, sample determination, saturation); and (3) how decision analysis theory and examples from existing practice were used to create a requisite structured approach for writing evaluation criteria.

The structured approach proposed in this research has the potential to enhance decisions in best-value procurement because this study recommends more precise and specific evaluation criteria that will strengthen evaluation results. This will allow agencies to more clearly understand the outcomes of their evaluation criteria. It will allow design-builders to understand what constitutes best-value and prepare better proposals. One of the major advantages of the proposed approach is that it provides a clear structure for the definition of evaluation criteria that will ensure internal validity and consistency in highway agencies decision-making.

CONCLUSIONS AND CONTRIBUTIONS

The analysis of current practice was taken from a broad set of design-build RFPs from across the US. In this data set, it was revealed that 43% of evaluation criteria are set using a general expression, which is neither direct scoring nor constructed-scale. In 53% cases, evaluation criteria are defined using a generic constructed scale, which does not allow measuring precisely the related requirements. Finally, in 4% of cases, evaluation criteria are specific and use direct scoring for assessing the requirements. Therefore, a high percentage of the evaluation criteria included in the current best-value request RFPs are imprecise and do not follow a structured formulation such as direct scoring or specific constructed-scale. This is a key limitation because, as it was shown in this research, the more generically defined an evaluation criterion is, the less capable of being comprehensive, direct, unambiguous, and understandable. These characteristics are, according to proven decision analysis theory, the desirable features of evaluation criteria.

Evaluation criteria not having these characteristics are not well suited to elicit meaningful information from the proposals to evaluate, compare, and select the best firm to develop the project. Further, they do not enable design-builders to know how highway agencies measure best-value. Thus, design-builders are more likely to fail in preparing proposals that offer the best-value that highway agencies need.

In order to help highway agencies to transition from current practice to more comprehensive, direct, unambiguous, and understandable evaluation criteria, this research provides a structured approach to guide the process of writing these evaluation criteria. The suggested approach comprises ten questions that relate to the selection objectives, the requirements, and the evaluation criteria characteristics. This approach guides the creation of consistent evaluation criteria that can be adapted to the various state highway agency evaluation criteria.

Using this approach might help to improve current practices. Improving current practices, in turn, might make a twofold contribution. First, consistent evaluation criteria can help decision-makers to strengthen their evaluation results. Bolstering evaluation criteria outcomes can help them to be more influential in the selection process. Second, consistent evaluation criteria can precisely show the design-builders what constitutes best-value and how best-value would be scored. Thus, design-builders would be able to prepare better proposals that offer the best-value required by the agencies. Overall, drafting more consistent evaluation criteria would enhance the best-value procurement as a whole.

This research contributes to the engineering management body of knowledge of alternative contracting methods procurement. To date, several authors have provided recommendations on how to write evaluation criteria (AASHTO 2018; AGC of America & NASFA 2008; DBIA 2019; Gransberg et al. 2006a; b; Molenaar and Tran 2015). However, these recommendations are broad, non-structured, and not based on a large sample analysis of current practice. This research contributes to this previous knowledge by providing a structured approach to writing evaluation criteria. This approach, unlike the previous recommendations, is based on proven decision analysis theory and the study of 540 best-value evaluation criteria used in the current practice of design-build RFPs.

This research has been developed based on the evaluation criteria information included in the current design-build RFPs. Generally, DOTs use internal procedures to evaluate the proposals that might consist of more detailed evaluation criteria. For this reason, future research should be conducted through DOTs case studies in order to check and complement these findings. Overall, these case studies could further be used to explore the efficacy of the proposed approach.

This research addresses how to better write evaluation criteria to select the proposer that offers the best-value. In best-value procurement, however, obtaining the best-value might depend not only on how to write evaluation criteria but also on what areas are assessed, what scoring practices and weights are used, and what award algorithm is implemented. How these elements might influence in obtaining best-value by highway agencies should be addressed by future research.

DATA AVAILABILITY STATEMENT

All data that support the findings of this study are available from the corresponding author upon reasonable 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 Pontifical Catholic University of Chile.

REFERENCES

AASHTO. (2018). *AASHTO Guide for Design-Build Procurement*. AASHTO, Washington D.C.

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

Bana, C. A., Correa, E. C., De Corte, J.-M., and Vansnick, J. C. (2002). “Facilitating bid evaluation in public call for tenders : a socio-technical approach.” *Omega. The International Journal of Management Science*, 30, 227–242.

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

Catanzaro, M. (1988). “Using qualitative analytical techniques. In Nursing Research; Theory and Practice.” *Nursing research theory and practice*, 437–456.

Chini, A., Ptschelinzew, L., Minchin, R. E., Zhang, Y., and Shah, D. (2018). “Industry Attitudes toward Alternative Contracting for Highway Construction in Florida.” *Journal of Management in Engineering*, 34(2), 1–16.

Choi, K., Jung, I., Yin, Y., Gurganus, C., and Jeong, H. D. (2020). “Holistic Performance Evaluation of Highway Design-Build Projects.” *Journal of Management in Engineering*, 36(4), 1–11.

Clemen, R. T., and Reilly, T. (2000). *Making Hard Decisions with Decision Tolls*. Cengage Learning, Mason, OH.

Collier, Z. A., and Lambert, J. H. (2019). “Principles and methods of model validation for model risk reduction.” *Environment Systems and Decisions*, Springer US, 39(2), 146–153.

DBIA. (2017). “2017 State Statute Report.” <https://dbia.org/wp-content/uploads/2018/06/2017-State-Statute-Report-DBIA.pdf>.

DBIA. (2019). “Position Statement. Principles of Best Value Selection Principles of Best Value Selection.” <https://store.dbia.org/product/dbia-position-statement-best-value-selection/> (Feb. 5, 2020).

De_Araújo, M. C. B., Alencar, L. H., and de Miranda Mota, C. M. (2017). “Project procurement management: A structured literature review.” *International Journal of Project Management*, 35(3), 353–377.

Dehghan, R., and Ruwnapura, J. Y. (2014). “Model of trade-off between overlapping and rework of design activities.” *Journal of Construction Engineering and Management*, 140(2), 1–13.

Elo, S., and Kyngäs, H. (2008). “The qualitative content analysis process.” *Journal of Advanced Nursing*, 62(1), 107–115.

FMI. (2018). “Design-Build Utilization. Combined Market Study.” <https://dbia.org/wp-content/uploads/2018/06/Design-Build-Market-Research-FMI-2018.pdf>.

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.

Gass, S. I. (1983). “Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis.” *Operations Research*, 31(4), 603–632.

Gransberg, D. D., Koch, J. E., and Molennar, K. R. (1986). “Writing Design-Build Performance Criteria-Three.” 67–89.

Gransberg, D. D., Koch, J. E., and Molennar, K. R. (2006). “Summary of Design-Build Contracting.” *Preparing for Design-Build Projects*, 251–255.

Greenberg, H. J. (1988). “Validation of Decision Support Systems.” *Mathematical Models for Decision Support*, G. Mitra, ed., Springer-Verlag, Berlin Heidelberg.

Howard, R. A. (1966). “Decision Analysis: Applied-Decision Theory.” *4th Internal Conf. on Operational Research*, Boston, Mass.

Howard, R. A. (1980). “An Assessment of Decision Analysis.” *Operations Research*, 28(1), 4–27.

Howard, R. A. (1983). “Decision Analysis in Systems Engineering.” *The Principles and Applications of Decision Analysis.*, Strategic Decision Group, 57.

Howard, R. A. (1988). “Decision Analysis: Practice and Promise.” *Management Science*, 34(6), 679–695.

Keeney, R. L., and Gregory, R. S. (2005). “Selecting Attributes to Measure the Achievement of Objectives.” *Operations Research*, 53(1), 1–11.

Keeney, R. L., and Raiffa, H. (1976). *Decisions with multiple objectives*. John Wiley & Sons., New York.

Kerzner, H. (2017). *Project Management Metrics, KPIs, and Dashboards: A Guide to Measuring and Monitoring Project Performance.* John Wiley & Sons.

Lucko, G., and Rojas, E. M. (2009). "Research Validation: Challenges and Opportunities in the Construction Domain." *Journal of Construction Engineering and Management*, 136(1), 127–135.

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.

Montalbán-Domingo, L., García-Segura, T., Amalia Sanz, M., and Pellicer, E. (2019). "Social Sustainability in Delivery and Procurement of Public Construction Contracts." *Journal of Management in Engineering*, 35(2), 1–11.

NCOITS (North Carolina Office of Information Technology Services). (2008). "Establishing Effective Evaluation Criteria and an Effective Scoring Method." <www.its.state.nc.us/itprocurement/>.

Papajohn, D., El Asmar, M., and Molenaar, K. R. (2019). "Contract Administration Tools for Design-Build and Construction Manager/General Contractor Highway Projects." *Journal of Management in Engineering*, 35(6).

Phillips, L. D. (1984). "A theory of requisite decision models." *Acta Psychologica*, 56(1–3), 29–48.

PMI. (2016). *Requirements Management. A Practice Guide*. Project Management Institute.

Porter-Roth, B. (2007). "Proposal evaluation criteria." *The AIIM guide to ECM purchasing*, 46–47.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., and Jinks, C. (2018). "Saturation in qualitative research: exploring its conceptualization and operationalization." *Qual Quant*, 52, 1893–1907.

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.

Shalwani, A., Lines, B. C., and Smithwick, J. B. (2019). "Differentiation of Evaluation Criteria in Design-Build and Construction Manager at Risk Procurements." *Journal of Management in Engineering*, 35(5), 1–9.

Tax Policy Center. (2017). "State & Local Government Finance Data Query System." <https://www.taxpolicycenter.org/statistics/state-and-local-general-expenditures-capita> (Mar. 3, 2020).

Touran, A., Gransberg, D. D., Molenaar, K. R., and Ghavamifar, K. (2011). "Selection of Project Delivery Method in

Transit: Drivers and Objectives.” *Journal of Management in Engineering*, 27(1), 21–27.

Tran, D. Q., Brihac, A., Nguyen, L. D., and Kwak, Y. H. (2018). “Project Cost Implications of Competitive Guaranteed Maximum Price Contracts.” *Journal of Management in Engineering*, 34(2), 1–11.

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

UTHealth (University of Texas Health Science Center). (2020). “Formal Bid Procedures.” <https://www.uth.edu/dotAsset/1a0208cb-418d-444d-8081-c6ddfea9824c.pdf>.

Vieira, A. C. L., Oliveira, M. D., and Bana, C. A. (2020). “Enhancing knowledge construction processes within multicriteria decision analysis : The Collaborative Value Modelling framework ☆.” *Omega. The International Journal of Management Science*, Elsevier Ltd, 94, 102047.

Winterfeldt, D. Von, and Fasolo, B. (2009). “Structuring decision problems : A case study and reflections for practitioners.” *European Journal of Operational Research*, Elsevier B.V., 199(3), 857–866.

Xia, B., Chan, A., Zuo, J., and Molenaar, K. (2011a). “Analysis of Selection Criteria for Design-Builders through the Analysis of Requests for Proposal.” *Journal of Management in Engineering*, 29(1), 19–24.

Xia, B., Skitmore, M., and Zuo, J. (2011b). “Evaluation of Design-Builder Qualifications through the Analysis of Requests for Qualifications.” *Journal of Management in Engineering*, 28(3), 348–351.

TABLES

Table 1 Research data

#	State	# Projects	# EC.	#	State	# Projects	# EC.
1	California	4	11	**12**	Virginia	4	16
2	Nueva York	5	38	**13**	Washington	10	33
3	Texas	5	26	**14**	Alaska[a]	1	1
4	Florida	8	34	**15**	Arizona	1	4
5	Ohio	2	55	**16**	Colorado	4	14
6	Georgia	3	12	**17**	Connecticut	2	23
7	Kentucky	3	47	**18**	Louisiana	6	62
8	Maryland	10	30	**19**	Mississippi	7	38
9	Michigan[a]	1	1	**20**	South Carolina	10	37
10	Minnesota	6	17	**21**	Utah	1	4
11	North Carolina	5	35		**# Total projects**	**98**	**540**

Note: [a] Formulation was taken from the state design-build manual

Table 2 Example of an evaluation criterion using a generic constructed scale defined in broad terms

Adjective Rating	Percentage of points awarded	Description
Excellent	90%-100%	The Proposal exceeds in a significant manner stated requirements in a beneficial way, providing advantages, benefits, or added value to the project, and provides a consistently outstanding level of quality.
Very good	80%-90%	The Proposal exceeds the stated requirements in a beneficial way, providing advantages, benefits or added value to the project, and offers a significantly better than acceptable quality.
Good	70%-80%	The Proposal comfortably meets the stated requirements, provides some advantages, benefits, or added value to the project and offers a generally better than acceptable quality.
Fair	50%-70%	Design-builder has demonstrated an approach that is considered to marginally meet stated requirements and meets a minimum level of quality.
Poor	0% (Failing)	Design-builder has demonstrated an approach that contains significant weaknesses/deficiencies and/or unacceptable quality.

Table 3 Example of evaluation criterion using a generic constructed scale and a prioritization of requirements

Adjective Rating	Description
Exceptional	The design-builder has demonstrated a complete understanding of the subject matter, and the Proposal advances the Project goals to an exceptional level. The Proposal communicates an outstanding commitment to quality by a highly skilled team in all aspects of the Work. The Proposal outlines a strong approach to mitigating project-specific risks and inspires confidence that all contract requirements will be met or exceeded. The Proposal contains significant strengths.
Good	The design-builder has demonstrated a strong understanding of the subject matter, and the proposal advances the Project goals to a high level. The Proposal communicates a commitment to quality by an experienced team in all aspects of the Work. The Proposal defines an approach to mitigating project-specific risks with little risk that the design-build would fail to meet the requirements of the contract. The Proposal contains strengths that outweigh weaknesses.

Acceptable	The design-builder has demonstrated an adequate understanding of the subject matter, and the Proposal meets the Project goals. The Proposal communicates a commitment to quality Work by a qualified team. Project-specific risks have been identified, and the design-builder has a reasonable probability of successfully completing the Work. The Proposal contains strengths that are offset by weaknesses.
Unacceptable	The design-builder has not demonstrated an understanding of the subject matter, and the Proposal presents an approach that does not address the goals of the project. The Proposal fails to meet stated requirements and/or lacks essential information. The commitment to quality is not adequate, with Work performed by unqualified or unproven teams. Project-specific risks are not addressed, and the Proposal generates little confidence that the project requirements can be met. The Proposal contains deficiencies, significant weaknesses, and minor strengths, if any.

Table 4 Points assignment per route

Roadway segments	**Quality Evaluation weighting (max points)**
NB Route 30 corridor	5
SB Route 32 corridor	5
From Route 17 EB Off-Ramp to Woodbury Commons Northern Entrance	3

Table 5 Points assignment per travel time.

	Proposed design solution performance. Travel time (min)			
	AM peak hour	PM peak hour	Sat MD peak hour	Total
Design-builder A	2.00	3.52	3.46	8.98
Design-builder B	2.35	3.50	3.46	9.31

Table 6 Safety. Adapted from MnDOT. Albertville Project (2018)

Requirements	**Generic expression**
• Provide a narrative describing the Design-builder's approach to both mitigating safety risks and proactively enhancing safety practices. • State the degree to which the approach includes final commitments and list any specific commitments in a table.	This approach will be evaluated based on the likelihood and degree to which the Design-builder's commitments will achieve the following: • Safe working conditions for Contractor and DOT employees on the Site during construction. • Safe conditions for pedestrians and other people accessing the Project Right of Way during construction other than those traveling by vehicle.

Table 7 Safety. Requirements and evaluation criterion re-formulated

Requirements (Reformulated)	Evaluation criterion (Reformulated): constructed scale
• Provide a list of 10 safety commitments related to risk mitigation. • Each commitment must specify: - Estimated impact on the final project cost and schedule on a qualitative or quantitative basis[b]. - Estimated impact on injury rates, near misses and event with significant injuries [b]*Quantitative evaluation of impacts is preferred to qualitative assessment.*	**Excellent.** The whole set of commitments demonstrates the highest[c] reduction in injury rates, near misses, and events with significant injuries. **Good.** The whole set of commitments demonstrates an average reduction in injury rates, near misses, and events with significant injuries. **Fair.** The whole set of commitments demonstrates the lowest reduction in injury rates, near misses, and events with significant injuries. **Poor.** None of the previous case Notes: [c]*Highest, average, lowest is defined based on all proposals impacts' values* *Within each category, proposals will be ranked based on their ability to minimize the effects on the overall project's cost and schedule.*

Table 8 Safety. Adapted from KYTC. Boone. Route I-275 Project (2019)

Requirements	Evaluation criterion formulation: Generic constructed scale	
	Scoring range	**Description**
• Describe the safety considerations specific for this project. • Discuss the firm's overall approach to safety	90-100	The Technical Proposal component demonstrates an approach that is considered to significantly exceed the ITP requirements and objectives beneficially (providing advantages, benefits, or added value to the project), and that provides a consistently outstanding level of quality. Must have a significant strength or number of strengths and no weaknesses.
	80-89	The Technical Proposal component demonstrates an approach that is considered to exceed the ITP requirements and objectives in a beneficial way (providing advantages, benefits, or added value to the project) and offers a generally better than acceptable quality. Must have strengths and no significant weaknesses.
	70-79	The Technical Proposal component demonstrates an approach that is considered to meet the ITP requirements and objectives and offers an acceptable level of quality. It has strengths, even though minor or moderate weaknesses exist.
	60-69	The Technical Proposal component demonstrates an approach that is marginally acceptable.
	0-59	The Technical Proposal component demonstrates an approach that contains no strengths and contains minor or significant weaknesses.

Table 9 Safety. Requirements and evaluation criterion re-formulated

Requirement (Reformulated)	Evaluation criterion formulation: Specific constructed-scale (Reformulated)	
	Scoring range	**Description**
• Describe the main 5 risks affecting safety that are specific for this project. Each one must include o Estimated probability of occurrence. Including rationale. o Impact on project cost and schedule. Including qualitative analysis. o Prevention and mitigation measures. Including associated costs	90-100	Proposals in this range identify the most relevant risks providing a comprehensive rational for their probability of occurrence and impact. These proposals also include both prevention and mitigation measures with the best cost-effectiveness ratio among the proposals.
	80-89	Proposals in this range identify the most relevant risks providing a comprehensive rational for their probability of occurrence and impact. These proposals also include both prevention and mitigation measures with an average cost-effectiveness ratio among the proposals..
	70-79	Proposals in this range identify the some of the relevant risks providing a rational for their probability of occurrence and impact. These proposals also include both prevention and mitigation measures with the best cost-effectiveness ratio among the proposals of this type
	60-69	Proposals in this range identify some of the relevant risks providing a rational for their probability of occurrence and impact. These proposals also include both prevention and mitigation measures with the average cost-effectiveness ratio among the proposals of this type
	0-59	None of the previous

Table 10. Maximize overall safety, capacity, and operation. Adapted from CDOT. Cimarron project (2014)

Requirements	Evaluation criterion formulation: Direct scoring	
Design-builders should submit the Additional Requested Elements (ARE) that they consider among the following proposed by this Agency:	Each ARE has different points associated as Table 7 shows:	
1. Full-Width I-25 Bridges	**Additional Requested Elements (ARE)**	**Points**
2A. Widen US 24 Bridge over Fountain Creek and provide Additional lanes to 8th	1. Full-Width I-25 Bridges	3
2B. Replace US 24 Bridge over Fountain Creek and provide Additional lanes to 8th	2A. Widen US 24 Bridge over Fountain Creek and provide Additional lanes to 8th	4
3. Trail and Creek Improvements along Fountain Creek up to 8th street	2B. Replace US 24 Bridge over Fountain Creek and provide Additional lanes to 8th	13
4. Contractor Defined ARE (additional operational Improvements on US 24 and at the I-25 and US 24 Interchange	3. Trail and Creek Improvements along Fountain Creek up to 8th street	2
	4. Contractor Defined ARE (additional operational Improvements on US 24 and at the I-25 and US 24 Interchange	7
	Maximum Subtotal points (ARE 2A and 2B are Mutually Exclusive)	**25**

	For AREs 1, 2, and 3, each ARE included in the Proposal will be given the total number of points available for that ARE. ARE 4 will be variably scored up to a maximum of 7 points based on its effectiveness at providing additional capacity and operations improvements on US 24 and at the I-25 and US 24 interchange

Table 11 Requirements and evaluation criterion re-formulated

Requirements (Reformulated for ARE 4)	Evaluation criterion formulation: Direct scoring (Reformulated for ARE 4)
4. Contractor Defined ARE (additional operational Improvements on US 24 and at the I-25 and US 24 Interchange including, for this solution • Reduction in waiting time in peak hours (RWT) • Average travel time saving per day (TTS) Data calculated using "X" traffic simulation software and comparing the scenarios with and without ARE 4.	ARE 4 will be variably scored up to a maximum of 7 points based on its capacity to minimize the waiting time in peak hours and the average travel time. The design-builder with the maximum RWT + TTS will be score 7 points. The remaining design-builders' time will be pro-rated against the maximum time.