PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

ADVANCING KNOWLEDGE OF DESIGN-BUILD AND BEST-VALUE IN HIGHWAY PROJECTS

MARIA INMACULADA CALAHORRA JIMÉNEZ

Thesis submitted to the Pontificia Universidad Católica de Chile and University of Colorado Boulder in partial fulfillment of the requirements to receive the degree of Doctor in Engineering Sciences and Doctor of Philosophy

Advisors:

LUIS FERNANDO ALARCÓN

ALONDRA CHAMORRO

KEITH MOLENAAR

CRISTINA TORRES-MACHI

Santiago de Chile, July, 2020

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

ADVANCING KNOWLEDGE OF DESIGN-BUILD AND BEST-VALUE IN HIGHWAY PROJECTS

MARIA INMACULADA CALAHORRA JIMÉNEZ

Members of the Committee:

LUIS FERNANDO ALARCÓN

ALONDRA CHAMORRO

KEITH MOLENAAR

CRISTINA TORRES-MACHI

PAZ ARROYO

HERNÁN DE SOLMINIHAC

PAUL GOODRUM

GUSTAVO LAGOS

Thesis submitted to the Pontificia Universidad Católica de Chile and University of Colorado Boulder in partial fulfillment of the requirements to receive the degree of Doctor in Engineering Sciences and Doctor of Philosophy

Santiago de Chile, July 2020

ADVANCING KNOWLEDGE OF DESIGN-BUILD AND BEST-VALUE IN HIGHWAY PROJECTS

by

MARIA CALAHORRA-JIMENEZ

B.S-M.S., University of Cantabria, Santander, 2004

M.S., University of Chile, Santiago de Chile, 2015

A thesis submitted to the

Faculty of Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental, and Architectural Engineering

2020

Committee Members:

Keith R. Molenaar

Cristina Torres-Machi

Paul Goodrum

Luis Fernando Alarcón

Alondra Chamorro

ABSTRACT

Calahorra-Jimenez, Maria (Ph.D., Civil Engineering, Department of Civil, Environmental, and Architectural Engineering)

Advancing Knowledge of Design-Build and Best-Value in Highway Projects.

Thesis directed by Professor Keith Molenaar, and Assistant Professor Cristina Torres-Machi

Highway agencies use contracting strategies such as design-bid-build (DBB), low-bid, design-build (DB), and best-value procurement to deliver their projects. Previous studies have shown that the use of DB and best-value helps improve cost project performance. However, some countries remain using DBB with consequent high-cost overruns, and other countries do use DB and best-value but fail in reaching their full potential. Thus, there is a need to revisit current DBB-low bid and DB-best-value practices and provide new knowledge to enhance their performance. To this end, this research aimed to (1) mitigate the main reasons for cost overruns in DBB low bid road projects by exploring measures based on the DB project delivery system; (2) enhance the influence of non-cost factors in DB best-value selection by proposing a systematic approach to write evaluation criteria, and (3) promote a better balance between cost and non-cost factors in DB best-value selection by proposing empirical-based ranges of weights to use in the best-value weighted criteria algorithm. These objectives were achieved by conducting three independent research. The first research provided a mapping of the primary reasons for cost overruns into the phases of design, procurement, and construction by examining 41 interviews using inductive content analysis. It determined what reasons were specific for each of the phases and discussed how the DB delivery system might help minimize these reasons. The second research identified how highway agencies articulated best-value evaluation criteria and proposed a structured approach to enhance current practice on writing these criteria. Through the lens of decision analysis, the researchers conducted a content analysis on 540 criteria included in 98 requests for

proposals (RFPs) from 21 states across the United States (US). The analysis found three levels of specificity in writing evaluation criteria. Building upon these levels and on decision analysis theory principles, the research proposed a structured approach to support highway agencies in the process of crafting evaluation criteria. The third research suggested a range of weights to use in best-value procurement, aiming to reach a better balance between cost and non-cost factors in the selection. The research characterized current scoring trends based on a unique data set of 882 non-cost scores and 1,148 cost scores from best-value procurement bidding results. These data were used to simulate the weighted criteria algorithm and explore the balance between cost and non-cost factors. As a result, recommendations on weights were provided to help highway agencies to implement a "real" best-value selection. This research contributes to the body of knowledge of cost overruns in DBB road projects and best-value procurement for DB highway projects. It is also a unique contribution to the improvement of current state-of-the-practice of agencies managing the design and construction of highway projects.

RESUMEN

En Chile, 74% de los proyectos viales experimentan sobrecostos. En EEUU, 80% de estos proyectos se adjudican a la empresa más barata. Lo anterior, en última instancia, repercute negativamente en la gestión de recursos públicos.

El objetivo de esta investigación fue desarrollar conocimiento para mitigar sobrecostos en proyectos viales de Chile; y hacer más influyentes los criterios técnicos en la selección de los equipos en EEUU.

En Chile, se entrevistaron 41 profesionales y se analizaron formas colaborativas de gestión de proyectos. En base a ello, se propusieron, como principales medidas de mitigación, la integración temprana del constructor y el uso de licitaciones basadas en el "mejor valor".

En EEUU, se utilizó teoría de análisis de decisiones y se estudiaron 540 criterios de evaluación para proponer un enfoque estructurado que guie la creación de criterios más influyentes. Además, el análisis de 1.970 puntuaciones técnicas y económicas y el uso de simulación, determinó que ponderaciones del costo mayores de 57%, implican selecciones que no considera criterios técnicos.

Esta investigación permitirá que los gestores públicos minimicen sobrecostos y elijan a los mejores equipos. Futura investigación mejorará las formas de contratación y desarrollo de proyectos como medio para mejorar el desempeño de proyectos viales.

Palabras Claves: Design-Build, Best-Value, Highways, Evaluation Criteria, Cost Overruns

DEDICATION

A mis padres Víctor y Maricarmen.

Vuestro ejemplo de esfuerzo y humildad siempre guía mis pasos.

A Gustavo.

Compañero de doctorado, compañero de vida.

Acknowledgments

This thesis has been a life-changing experience, and I am very grateful to so many people that have helped me in this accomplishment. I will start with my advisors. Professor Luis Fernando Alarcon was the one that opened me the doors to academia. After working for 14 years in the industry, I was completely new in this world. Luis Fernando trusted me, helped me to grow (with him, I awarded my first research project!), and introduced me to Lean (I did not know anything!). Thank you, Luis Fernando, for everything. Professor Alondra Chamorro was the one that always encouraged me to be rigorous with my research methodologies. She has always had a constructive criticism that has helped me to improve a lot my research skills. Thank you, Alondra, for all your advice.

Professor Keith Molenaar has been a "milestone person" in my life. I do not have words to say how grateful I am for all the opportunities he has given me. Opportunities that have changed my life completely. Welcoming me to the US., considering me as part of the SCDOT research project, always encouraging, always providing sound advice, always being there. Thank you, Keith, for being the great person that you are.

Last but not least, Professor Cristina Torres-Machi. Cristina has been my greatest support in the development of the thesis. She has always helped me deal with all types of challenges. Re-thinking different approaches the number of times that was needed (she knows that I am a little bit stubborn!). Providing me quick feedback so I could progress at a good pace. Guiding me in all types of Ph.D. processes. I could keep writing, but the main thing here, Cristina, is that I want to acknowledge your excellent quality as an advisor. Thank you very much for really advise me, and help me to grow in all this path.

I feel very fortunate about having the four of you as advisors. I have had the opportunity of learning from your different knowledge and approaches, and this is something that I will always keep with me.

I also want to acknowledge the time and the excellent feedback of all the members of my committee. Professor Paz Arroyo, Professor Hernan de Solminihac, Professor Gustavo Lagos, and Professor Paul Goodrum. All your comments in the candidature exam, comprehensive exam, and follow-up meetings really served me to think my work under different approaches, which is an enriching experience. This thesis has a little bit of all of you in it. Thank you.

I want to thank the Vicerrectoria de Investigacion and the Escuela de Ingenieria de la Pontificia Universidad Católica de Chile for the economic support provided. Without it, it would have been much more difficult to keep studying in the US when I quit my job in Chile. I really appreciate this help. Thank you very much.

Also, I am very grateful for the opportunity of working for the South Carolina Department of Transportation (SCDOT). Chris Gaskins is doing a fantastic job leading the SCDOT Design-Build highway program. Thank you for having me on the team.

CU Boulder has excellent programs and resources to succeed as a Graduate Student. I especially appreciate all the support from the Graduate School and specifically from Sara Tynen. Sara leads the "Publication Bootcamp" and the "3 Minute Thesis" superbly. These activities were extremely useful for improving my writing and public speaking skills. Also, and with the Covid-19 pandemic, I have felt that Sara is always there in case it is needed. Thank you very much, Sarah!

I want to make another special mention for the "Clearly Speaking" program from the Speech, Hearing and Language Clinic at CU Boulder. Thank you very much for all the excellent professionals that have helped me to improve my English pronunciation. Samantha, Alison, Jenny, Caitlin, Kate, and Alison Lemke: thank you! I can't believe how Gustavo and I have improved thanks to you. Also, I would like to acknowledge Kristen Drybeard. My great writing advisor from the Writing Center. Thanks to her, I overcame my fear of writing scientific articles in English. It was terrific having your guide and see how I have improved! (Although improving writing never ends!). Thank you very much, Kristen!

In this process, I have found great colleagues with whom I share the common aim of improving society with our research. In Chile, Rodrigo, Luis, Johnathan, Zulay and Roberto. In Boulder, Desiree, Scot, Adam, Kelly, Doug, Suraj, Alfonso, Elif, Tejas, Mel, Bil, Bashar, Vishwa, Kimmy, Briar, Janice, Guillermo, Sarah and Victor. Special mention to Harrison Mesa, who shared with me his experiences and knowledge before coming to Boulder; and, Laura Montalban, for our talks and your help with my first research. For your support and your advice, thank you very much, Harrison and Laura.

I also want to acknowledge my "close-to-heart-team" who always accompanied me wherever I am, whatever I do. My parents, Victor y Maricarmen, my sister, Vir, my brother, Edu, my aunts Sonsoles and Teresa and my close friends "Churras y Merinas": Bea, Tery, Irene, and David.

Finally, Gustavo, thank you. Thank you for being with me on this journey. From that day in Chile, looking at potential Ph.D. programs (when I had no idea what to choose) to today. Yes, from that day to today, with all the extraordinary things that we have lived together between now and then. Thank you.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

"The real voyage of discovery consists, not in seeking
new landscapes, but in having new eyes."
—Marcel Proust

1.1 OBSERVED PROBLEM AND RESEARCH OPPORTUNITY

Traditionally, highway agencies have used design-bid-build (DBB) and low-bid as the contracting strategy for delivering highway design and construction projects. Through this strategy, the design and construction are hired independently, and teams are selected only on the basis of cost. However, two decades ago, highway agencies began to use alternative contracting strategies such as design-build (DB) and best-value. These strategies enable the collaboration between designers and constructors and consider other factors in addition to cost when selecting the teams. Highway agencies started the transition to DB and best-value to overcome some of the drawbacks of the traditional approach.

DB and best value procurement have demonstrated several advantages as compared with the traditional approach. For example, DB allows more flexibility and collaboration than DBB in the development of the design and the procurement of a contractor for a given project (Sullivan et al. 2017). Further, previous research shows that, generally, DB performs better than DBB regarding cost growth and schedule growth (FHWA 2006a; Minchin Jr et al. 2013; Shrestha et al. 2007, 2012; Sullivan et al. 2017; Tran et al. 2018b). Best-value, on the other hand, has enabled highway agencies to consider not only cost (as in low-bid) but also other different factors in evaluating and selecting the best firm to award highway contracts. By using best-value, instead of low bid, highway agencies increase the value-added to the project for each dollar spent (Abdelrahman et al. 2008a).

Regardless of the advantages that DB and best-value might provide, some countries remain using the traditional approach. Other countries do use DB and best-value but fail to implement them to reach their full potential. For example, in Chile, DBB is the standard delivery system used in road projects. However, 74% of projects procured through DBB experienced cost overruns, and 71% of such projects experienced delays (MOP Chile 2016). Further, Tala (2015) reported that 70% of the reasons leading to cost overruns in Chilean public infrastructure projects related to problems from the pre-design and design phases that have repercussions into the construction phase.

The United States (US), on the other hand, began the use of alternative contracting strategies in the late 1990s (FHWA 2017a). However, current practice on DB highway projects shows that even when using best-value procurement, agencies are more influenced by cost than by any other non-cost factors when selecting the company to award the contract for a project (FMI 2018a; Gaikwad 2019; Hill International 2016). Therefore, there is an illusion of doing best-value because, in the end, the only thing that matters is cost, like in low-bid procurement.

These examples show a need to revisit current DBB-low bid and DB-best-value practices and provide new knowledge to enhance their performance. In Chile, DBB-low bid requires minimizing cost overruns. In the US, DB-best-value requires a reduction in the tendency of cost alone to drive project selection.

Minimizing cost overruns is preceded by the identification of the reasons that lead to these cost overruns. Previous research has addressed the identification of reasons leading to cost overruns worldwide (Adam et al. 2017; Al-Hazim et al. 2017; Arditi et al. 1985; Baloi and Price 2003; Bhargava et al. 2010; Creedy et al. 2010a; Khabisi et al. 2016; Rosenfeld 2014; Shane et al. 2009; Sohu et al. 2017; World Bank 2014). However, these studies usually resulted in lists and rankings of reasons for cost overruns in construction projects. They did not look at the phases in which those reasons occurred. Further, they did not explore how (since alternative delivery methods such as DB are better in controlling cost growth) strategies associated with alternative project delivery methods might help minimize those reasons for cost overruns.

Best-value procurement, on the other hand, has been profusely analyzed by previous research. For example, some studies have taken a general approach by summarizing procedures and best practices that served to create frameworks and recommendations for the best use of the method (Anderson and Russell 2001; Molenaar and Johnson 2003a; Molenaar and Tran 2015; Scott et al. 2006; Tran et al. 2017a). Other studies have focused on identifying the most frequent evaluation criteria used in best-value procurement (De Araújo et al. 2017; Gransberg et al. 1986; Scott et al. 2006; Shalwani et al. 2019; Xia et al. 2011a; b). Further, previous research has also focused on more specific elements of best-value procurement. For example, Scott et al. (2006) identified and analyzed the different scoring systems and award algorithms in use. Asmar et al. (2010) explored how subjectivity in scoring and weights might be removed using normalization and graphical models. Ballesteros-Perez et al. (2016) studied how weights and scores can influence the proposers' behavior. Schöttle et al. (2017) compared best-value selection against other methods to illustrate their different impacts on the tendering procedure. All these studies have contributed to a better understanding of the best-value selection. They have not explored, however, actual procurement documents and bidding results in order to identify the reasons leading to the current cost-driven tendency of best-value selection.

This thesis revisits current DBB-low bid and DB-best-value practices and aims to provide knowledge to enhance their performance. The research fills the gaps found in the identification and minimization of cost overruns in DBB-low bid highway projects. Further, the research addresses the existing gap in the analysis of the cost-driven tendency of best-value procurement in highway projects. The thesis filled the identified gaps through three independent research. In Chapter 2, the study addressed the reasons for cost overruns in DBB-low bid road projects. It diagnosed the main reasons for cost overruns in the design, the procurement, and the construction phases of DBB road projects. Based on them, the study pointed out the specific elements of DB project delivery that might help to minimize those reasons.

In Chapters 3 and 4, the thesis focused the analysis on investigating best-value procurement under two approaches. First (Chapter 3), the research explored best-value evaluation criteria and how their formulation might be improved in order to make non-cost evaluation criteria more influential in the selection. Second (Chapter 4), the research analyzed a large set of cost and non-cost scores to understand how the weight of cost and the ranges of scores influence the balance between cost and non-cost factors when using the weighted criteria award algorithm.

This research contributes to the body of knowledge of cost overruns in design-bid-build highway projects and best-value procurement for design-build highway projects. This study also is a unique contribution to the improvement of current state-of-the-practice of highway agencies managing the design and the construction of highway projects.

1.2 Background

According to the International Transport Forum, Organization for Economic Cooperation and Development (OECD) countries invest, on average, approximately 1% of their Gross Domestic Products on road and rail infrastructure construction and maintenance (OECD/ITF 2013). Thus, a considerable amount of tax-payer money is spent on infrastructure projects. However, the construction industry—that makes these projects real—" holds the dubious honor of having the lowest productivity gains of any industry" (The Economist 2017).

Research shows that reasons such as errors and omissions in design, lack of constructability or errors in information when conveying the design to actual construction are some of the main factors leading to cost overruns in construction projects worldwide (Creedy et al. 2010a; Forbes and Ahmed 2011; Khabisi et al. 2016; Shane et al. 2009; World Bank 2014). Further, an inappropriate selection of design and construction teams have shown to reduce the ability to deliver the projects on-budget (Amoatey and Ankrah 2017; Khabisi et al. 2016; Mevada and Devkar 2017; Morledge and Smith 2013).

Thus, both how highway agencies deliver the design and the construction phases (i.e., project delivery system) and how they select the design and construction teams (i.e., the procurement method) influence the project's performance in terms of cost, schedule, conflicts, claims, and disputes (Hashem M. Mehany et al. 2018; Mehany et al. 2017; El Wardani et al. 2006). It is essential, therefore, to consider which method will best balance the control of project cost and risk against achieving project objectives and outcomes (Commonwealth of Australia 2008).

1.2.1 Project Delivery Systems

The project delivery system is "the process by which a construction project is comprehensively designed and constructed including project scope definition, organization of designers, constructors and various consultants, sequencing of design and construction operations, execution of design and construction, and closeout and start-up" (Molenaar et al. 2014).

The project delivery system that public agencies have traditionally used to manage the phases of design and construction might influence the likelihood of having project performance problems. The traditional design-bid-build delivery system (DBB) involves contracting the design and the construction of a project independently. This traditional method has been used for a long time because it provides some significant advantages. It has a long history of acceptance. It offers open competition, a clear distinction between roles, and also, it allows easiness to develop the process of bidding (Pakkala 2002; Scott et al. 2006). This traditional delivery system, on the other hand, generates fragmentation, silos of knowledge and expertise, and a unilateral effort by the parts involved (AIA 2007).

Some of the flaws inherent in DBB have been addressed using other methods, such as Design-Build (DB). In this approach, public administrators procure together both the design and construction (Miller 2000). In a given project, DB allows more flexibility and collaboration in the development of the design and the procurement of a contractor. Although this method has faced cultural and legislative barriers in its implementation, in recent years, it has become a viable option for public construction projects (Sullivan et al. 2017).

As DB has been implemented, numerous studies regarding its effectiveness compared to DBB have been developed. Table 1 shows some of the research that has compared cost and schedule performance between DBB and DB projects.

Table 1 Studies regarding the effectiveness of DBB and DB project delivery methods.

Author(s)	Cost growth	Schedule growth
Sanvido and Konchar (1999)	DB projects show a greater chance of finishing within 5% of the budget.	DB more certainty in finishing on time.
FHWA(2006a)	DBB: 8.1%; DB: 4.3%	DBB: 11.6%; DB: -1.2%
Shrestha et al. (2007)	DBB: 4.12%; DB: -5.47% Statistically significant difference (alpha 0.05).	DBB: 12.88%; DB: 7.59% No significant difference (alpha 0.05).
Shrestha et al. (2012)	DBB: 6.3%; DB: 7.8% No significant difference (Alfa level 0.05).	DBB: 5.1%; DB: 20.5%. No significant difference (level 0.05).
Minchin Jr et al. (2013)	DBB projects performed significantly better regarding cost according to all statistical and arithmetic tests.	DBB projects did not compare as favorably regarding the duration
Plusquellec et al. (2017)	The results of the literature review show that DB outperforms DBB regarding cost and schedule growth.	
Tran et al. (2018b)	The results show that DB projects statistically perform better than DBB regarding cost growth and schedule growth for different construction and 3R* project work types.	

*Resurfacing, restoration, rehabilitation

Roughly 70% of the studies included in Table 1 state that the DB method performs better than DBB regarding cost growth. In comparison, 80% of the studies agree that the DB method outperforms DBB in terms of minimizing schedule growth.

In the USA, DB is continually growing as a delivery method in highway projects. During 2013-2017, the percentage of use of DB in these types of projects was 35%. This percentage is expected to grow to up to 43% between 2018 and 2021 (FMI 2018a). DOTs are embracing DB because it contributes to multiple benefits such as saving construction cost, improved quality, and benefits of innovative solutions (DBIA 2014).

On the contrary, in Chile, DBB is the standard delivery system used in road projects. DBB might be suitable for typical and common projects. However, projects involving high risk and many unknowns, as well as projects with a constrained schedule, require another type of delivery method (Molenaar et al. 2014). Highway agencies should be aware of the limitations of DBB and do not get carried away by a false feeling of saving money. For example, in Chile, according to a 2016 study, 74% of projects procured through DBB experienced cost overruns, and 71% of such projects experienced delays (MOP Chile 2016). Therefore, for those countries where DBB is the only option, it is relevant to establish bridges for transitioning to more collaborative possibilities for delivering highway projects.

Overall, alternative project delivery systems have shown to improve project performance as compared with the traditional DBB. Therefore, the analysis of the reasons for cost overruns in road projects under the lens of these alternative project delivery might help to find mitigation measures to reduce the main reasons leading to cost overruns. Further, this approach will serve highway agencies restricted to only use DBB as a first stepping stone to transition toward alternative delivery methods.

1.2.2 PROCUREMENT METHODS

The procurement involves acquiring goods, services, or construction through a third party at the moment that they are needed, at the quality that is prescribed, and at the best price possible. (Ruparathna et al. 2015).

The low bid is the traditional procurement type used in DBB. Through low bid, contractors are selected solely on the basis of price. With this method, the bidding process and selection criteria are transparent, and the construction companies that submit proposals are less likely to complain if their proposals are not selected. The shortcoming of this method is that it assumes that plans and specifications will be free of error, and it presumes that minimum requirements will meet the client's needs. Further, this method does not consider quality factors in the selection (Gransberg and Ellicott 1996).

Already in 1998, Sir John Egan (1998) stated that the trend of procuring design and construction teams based entirely on the tendered price was one of the most significant obstacles to meet expected project goals. Thus, he claimed that when selecting designers and constructors, public administrations needed to consider the difference between the lowest price and best value (Egan 1998). Abdelrahman et al. (2008a) added that the objective of transferring from the low bid to best-value procurement is to increase the value-added to the project for each dollar spent.

Best-value procurement, in contrast to low-bid, considers other factors in addition to costs when evaluating contractors who have placed bids. It constitutes a multicriteria decision-making method (MCDM) because it aims to select the best alternative among different proposals on the basis of a set of decision criteria (Triantaphyllou 2000).

Previous research has proposed different 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, highway agencies use easy-to-implement methods such as adjusted score (i.e., by multiplying the non-cost score by the estimated project price and dividing by the price proposal), adjusted bid (i.e., by dividing bid price by the non-cost score) and weighted criteria (i.e., weighted sum). From these methods, the weighted criteria is the most intuitive and transparent because of how the evaluation criteria are weighted and scored (Molenaar and Tran 2015).

Previous research on contractor selection have proposed new methodologies to evaluate and select the contractors but have not focused on methods that highway agencies are already using. Methods that, in practice, are not reaching their full potential. For example, the weighted sum method used in best-value procurement has not been addressed by this previous research on contractor selection.

The use of the weighted criteria best-value procurement comprises three main elements: evaluation criteria, weights, and scores. Evaluation criteria include cost and non-cost factors that highway agencies select in alignment with each project's goals. Weights represent the relative importance of each related criterion. Using scores, highway agencies rate the proposals in each of the evaluation criteria defined. Through the weight criteria algorithm, highway agencies combine criteria's weights and scores in order to obtain the award score, base on which the best proposer will be selected (Scott et al. 2006).

Evaluation criteria

Previous research has analyzed best-value evaluation criteria in design-build projects in terms of what evaluation criteria are most used in the request for qualifications and proposals (RFQ and RFP). For example, Xia et al. (2011b) performed a content analysis of 97 RFQ for public projects advertised between 2000 and 2011 in the United States. These authors identified 39 qualification criteria and classified them into the following eight categories: 1) experience, 2) project understanding and approach, 3) organizational structure and capacity, 4) past performance record, 5) professional qualifications, 6) responsiveness to RFQ, 7) Office location and familiarity with the local environment, and 8) legal status. Also, Xia et al. (2011a) analyzed the evaluation criteria included in 94 RFP from DB projects advertised between 2000 and 2010. These authors concluded that the following ten categories are the most commonly used: price, experience, technical approach, management approach, qualification, schedule, past performance, financial capability, responsiveness to the RFP, and legal status.

Overall, the results of these studies are beneficial because they enable us to identify the standard practice on best-value evaluation criteria. They are, however, not sufficient because they do not provide insight into the effectiveness of their use in the multicriteria selection process.

Weights and Scores

Among all the different contractor selection methods, this research focuses on the analysis of the weighted criteria award algorithm used in best-value procurement. This method is within the standard practices of highway agencies in the United States, and the aim is to provide knowledge to enhance its performance.

Previous research on best-value procurement over time have summarized procedures and best practices that served to create frameworks and recommendations for the best use of the method. For example, Molenaar and Johnson (2003b) surveyed the design-build two phases selection process focusing on the appropriate prequalification criteria as well as the amount of design that should be incorporated in the Request for Proposals (RFP). Further, Molenaar and Tran (2015) documented the state of practice for developing transparent best-value procedures. Tran et al. (2017a) investigated procurement and practice to promote transparency to improve the method. Scott et al. (2006) investigated best-value concepts in use at the moment of the research.

Overall, the results of these studies are beneficial because they enable us to identify the standard practice on best-value procurement. They are, however, not sufficient because they do not provide insight into the effectiveness of the performance of this decision-making method.

1.3 Intuition (Hypothesis)

This research poses the following two intuitions:

First, in the DBB delivery method, the segregation of the design and the construction phases and the use of low bid to procure the construction are source factors that facilitate several other reasons for cost overruns to occur. Previous research has shown that DB is an alternative project delivery that is better at controlling cost growth than DBB. DBB and DB are project delivery systems that manage the design, the procurement, and the construction phases of the projects differently. Thus, I posit that by mapping the reasons for cost overruns into the phases of DBB projects; I can find mitigation measures for these reasons based on DB project delivery characteristics.

Second, best-value procurement aims to achieve a best-value selection by balancing cost and non-cost factors in the evaluation of the proposals. To develop this assessment, highway agencies use different award algorithms. From those used in current practice, the weighted criteria algorithm is the most intuitive and transparent because of how the evaluation criteria are weighted and scored (Molenaar and Tran 2015). However, highway agencies need guidance on the use of this type of algorithm in order to achieve a best-value selection. I posit that by (a) defining how to articulate evaluation criteria and (b) determining what specific ranges for weights and scores should be used; I can improve the likelihood of having a better balance between cost and non-cost factors in the best-value selection.

1.4 Points of Departure

There are several gaps in our understanding of how contracting strategies in road and highway projects should be improved to reach a better application that leads to improved project performance

First, during the last two decades, alternative project delivery systems have served highway agencies to improve the control of cost growth (Sullivan et al. 2017). However, there is no research addressing the factors that make these alternative project delivery systems better than the traditional design-bid-build (DBB) in controlling cost growth. The identification of the reasons leading to cost growth, on the other hand, has been addressed profusely. Generally, these studies identify and prioritize the reasons for cost overruns. Still, they relate these reasons neither with the project delivery system nor with the phases that the project delivery system manages (the design, the procurement, and the construction).

In Chile, Alarcon et al. (2018) established recommendations to address the reasons for cost overruns based on the Lean paradigm. However, they did not discuss potential intervention measures based on other alternative project delivery methods more frequently used in highway practice, such as DB.

Thus, a research opportunity exists (Table 2) in identifying the reasons for cost overruns not aiming to list or rank but mapping them into the phases of design, the procurement, and the construction in order to find mitigation measures based on alternative project delivery systems such as DB.

Table 2 Points of Departure. Research 1.

Previous work		Opportunities
Previous studies identified the reasons for cost overruns in construction projects in the form of lists and rankings.	(Adam et al. 2017; Al-Hazim et al. 2017; Arditi et al. 1985; Baloi and Price 2003; Bhargava et al. 2010; Creedy et al. 2010a; Khabisi et al. 2016; Rosenfeld 2014; Shane et al. 2009; Sohu et al. 2017; Tala González 2015; World Bank 2014)	Identify the reasons for cost overruns by mapping these reasons in the design, the procurement, and the construction phases.
Previous study proposed Lean paradigm techniques to minimize the reasons for cost overruns in DBB road projects	(Alarcón et al. 2018)	Deepen the study of potential cost overruns mitigation measures based on alternative project delivery systems such as DB.

Second, best value procurement has been increasingly used by highway agencies in order to "select the contractor with the optimal combination of price and other factors to achieve project goals and improve the long-term performance" (Tran et al. 2017a). However, current practice shows that even when best-value procurement is used, the selection is driven more by cost than by any other factor (FMI 2018a; Gaikwad 2019; Hill International 2016).

Thus, a research opportunity exists (Table 3 and Table 4) in improving current best-value procurement in order to conduct a team selection process that better balances cost and non-cost factors. I addressed this research opportunity under two approaches. These approaches build upon the core elements of best-value procurement: (a) the evaluation criteria (Table 3), and (b) the weights and scores (Table 4).

Evaluation criteria. Decision-making theory establishes the principles that should rule the formulation of evaluation criteria in any decision-making process. However, previous research did not address how this theory applies to current practice in best-value decision-making. Generally, previous studies on best-value evaluation criteria focused on summarizing current practice. However, they did not analyze whether the formulation of evaluation criteria used in practice can be improved to make non-cost evaluation criteria more influential in the selection. Thus, the first approach for addressing the gap of improving current best-value procurement practices focuses on analyzing best-value evaluation criteria under the lens of decision-theory principles

Table 3 Points of Departure. Research 2.

Previous work		**Opportunities**
Theoretical study of the evaluation criteria in any decision-making process	(Belton and Stewart 2002; Clemen and Reilly 2000; Keeney and Gregory 2005; Keeney and Raiffa 1976)	Application of decision-making theory in the current practice of best-value procurement
General guidelines on best-value procurement transparency practices	(Molenaar and Johnson 2003b; Molenaar and Tran 2015; Scott et al. 2006; Tran et al. 2017a)	Specific guidelines on how to write evaluation criteria that warranty transparency of the evaluation process
Study of evaluation criteria in terms of **what** these evaluation criteria measure	(De Araújo et al. 2017; Xia et al. 2011a; b)	Study of best-value evaluation criteria in terms of **how** these evaluation criteria measure

Weights and scores. Previous research proposed different decision-making methods to select contractors. However, they did not focus on the improvement of a decision-making method currently in use. Further, previous studies analyzed and summarized current practice and proposed measures to improve the transparency of the method. However, they did no deepen the analysis of the specific values that the procurement method should use to balance cost and non-cost factors in the selection.

Thus, the second approach for addressing the gap of improving current best-value procurement practices focuses on determining ranges of weights and recommendations on scores that should be considered in best-value procurement to obtain a better balance between cost and non-cost factors.

Table 4 Points of Departure. Research 3.

Previous work		Opportunities
Proposal and summary of different methods and algorithms to select contractors	(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)	Analysis of the weighted sum algorithm used in best-value procurement aiming to reach a selection based on balanced criteria.
	(Molenaar and Johnson 2003b; Molenaar and Tran 2015; Scott et al. 2006; Tran et al. 2017)	
Theoretical approach to the analysis of weights and scores	(Barron and Barrett 1996) (Saaty 1980) (Lorentziadis 2010) (Asmar et al. 2010) (Ballesteros-pérez et al. 2016) (Ballesteros-pérez et al. 2015) (González-cruz 2012)	Analysis of historical weights and scores to explore the balance between cost and non-cost factors in the best-value selection

1.5 Research Questions

As a result of the problems observed, the intuitions and the theoretical points of departure, I formulate the following research questions:

Question 1. (Q1) How the main reasons for cost overruns in the design, procurement, and the construction phases of DBB road projects can be mitigated using DB strategies?

Question 2. (Q2) How evaluation criteria should be formulated to enhance the influence of non-cost factors in the best-value selection?

Question 3. (Q3) What ranges of weights and scores can better balance cost and non-cost evaluation criteria in the best-value selection?

1.6 RESEARCH METHODOLOGY

This dissertation ultimately aims to push the traditional DBB project delivery toward more collaborative approaches as well as to make best-value procurement reach its full potential as a decision-making method.

The study built the research methodology to answer the research questions that address the problems observed and consequent intuitions.

This dissertation was developed under the framework of the "Agreement for the joint supervision of Doctoral Thesis" between the Pontificia Universidad Católica de Chile and the University of Colorado Boulder. Therefore, the research uses data from both contexts. The study collected the Chilean data as part of the research project "Oportunidades para el Mejoramiento de la Gestión de Proyectos de Infraestructura Pública en Chile" (Alarcón et al. 2018) promoted by the Catholic University Public Center and developed in 2017. Further, the study gathered the data from the US's projects in the framework of the research project "Efficiency Study of SCDOT Design-Build Program" promoted by the South Carolina Department of Transportation. Finally, the research analyzed public data published and provided by the Departments of Transportation in the US related to the procurement of DB highway projects.

The overall methodology for this research is shown in Figure 1. This methodology displays the different steps under the framework of the three research questions. For each research question depicted on the left, research tasks and research methods are shown.

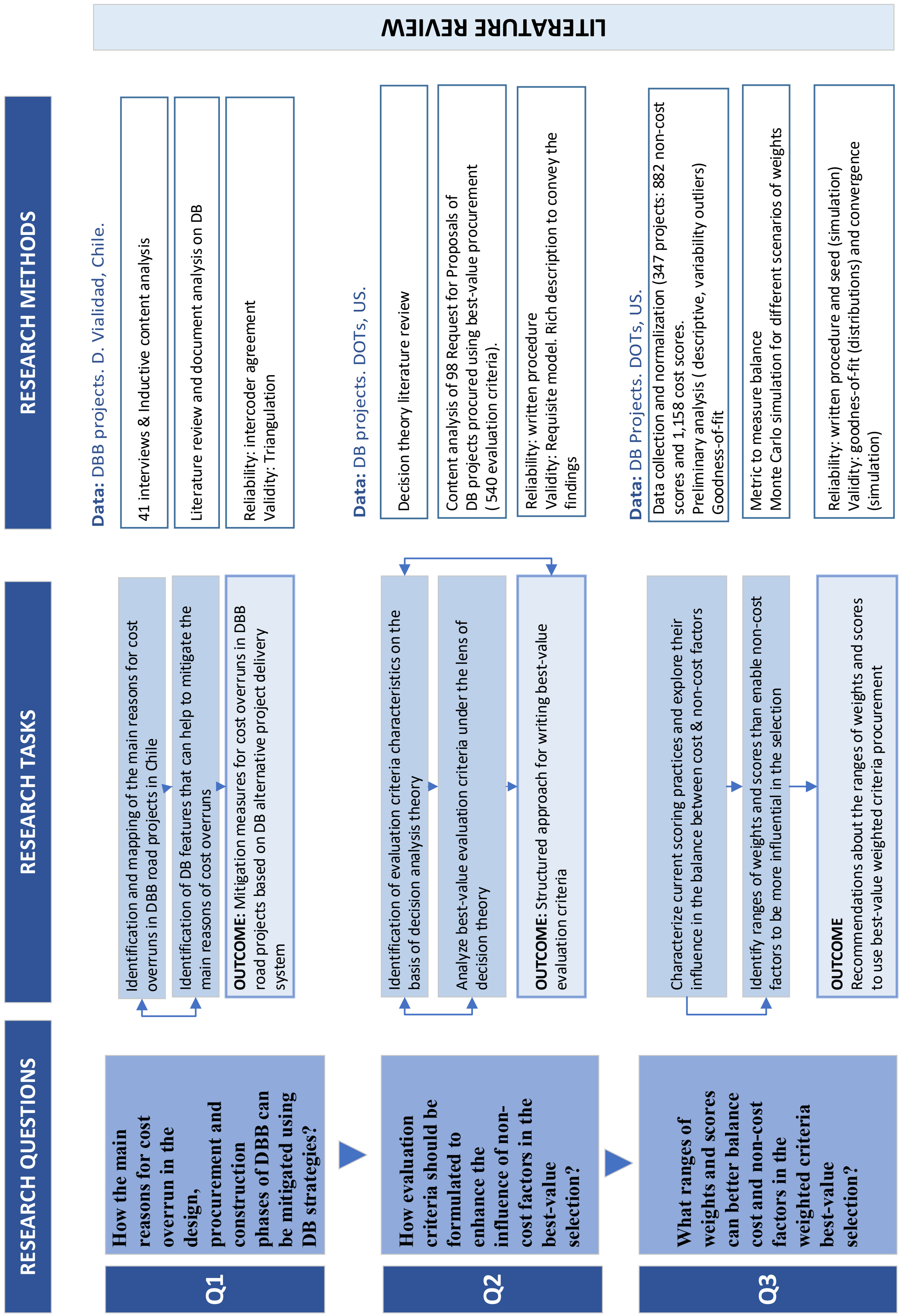


Figure 1 Thesis Methodology

This research was conceived under one philosophical worldview that underlies the research approach taken in each of the questions. According to Guba (1990), a worldview is a basic set of beliefs that guide action. In this dissertation, I considered a constructivist approach.

Under a constructivist worldview, "the researcher intends to make sense of (or interpret) the meanings other have about the world. Rather than starting with a theory (as in post-positivism), inquirers generate or inductively develop a theory or pattern of meaning" (Creswell 2009).

According to Creswell (2007), some of the relevant characteristics of a qualitative approach are

- Researches collect the data in the field where the problem under study happens.
- Researchers are the ones that collect the information.
- Use of multiple sources of data, such as interviews, observation, and documents.
- Researchers build their patterns and categories from the bottom up, by structuring data into increasingly more abstract information.
- In the whole process, researchers are focus on learning the meanings of the data they have collected.
- And, the initial plan of the research cannot be fixed, and researchers usually change them along the process.

The main idea behind this type of research is to learn about the problem from the information gathered and to address the research to obtain that information. (7). Researchers usually use a lens to view their studies. (8) Researchers interpret what they see and understand. (8) Researchers try to build a complex picture of the problem under study.

Question 1 was addressed using a qualitative approach because the identification of the reasons for cost overruns in DBB road projects is a context-based problem. It was possible to better understand the problem by gathering information from the professionals involved in the management of DBB road projects. Further, the study aimed to analyze the reasons for cost overruns under the lens of an alternative project delivery system (i.e., design-build) because these types of systems have shown to better control cost growth than DBB.

Question 2 was also addressed with a qualitative approach because there was a need to find meaning on how best-value evaluation criteria are currently written. The study used the lens of decision theory to detect to what extent the formulation of the evaluation criteria that highway agencies are currently using follow decision theory principles.

It is important to highlight that the qualitative approach of research 1 and 2 lead to a different concept of results' generalization. The aim of qualitative research is not to generalize findings to individuals, sites, or places outside those under study (Gibbs 2007). According to Creswell (2009), "the value of qualitative research lies in the particular description and themes developed in the context of a specific site." For this reason, research 1 and 2 do not include any statistical inferential space beside the one that is represented by the data sample in each case.

Finally, a quantitative-exploratory approach was considered for question 3 because the study aimed to explore the current patterns of scoring in best-value weighted criteria award algorithm and analyze how these patterns influence the balance between cost and non-cost evaluation criteria.

The following chapters describe in more depth the research approach for each of the questions.

1.7 Dissertation Organization

The dissertation is organized in a three-paper format. Each paper addresses one of the three research questions (Figure 2).

Chapter 1 introduces the problem, the background, points of departure and overall research methodology.

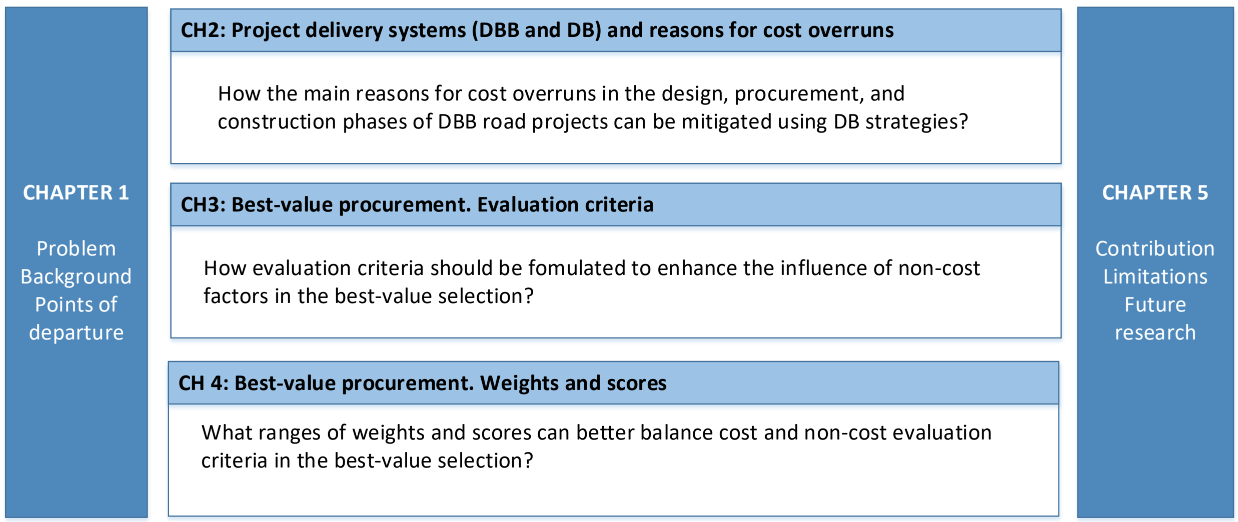


Figure 2 Dissertation organization

Chapter 2 (paper 1) is a manuscript accepted for publication in the Journal of Construction. This paper explores the reasons for cost overruns in highway projects and maps them into the phases of design, procurement, and construction. This paper contributes to a better understanding of the reasons for cost overruns and to how they could be mitigated through alternative project delivery systems.

Chapter 3 (paper 2) is a manuscript that is under only editor review in the ASCE Journal of Management in Engineering. This paper identifies the main patterns for writing evaluation criteria in the current practice of best-value procurement and proposes a structured approach to write them based on decision-theory. This research contributes to the improvement of agency practice in the authoring of best-value criteria.

Chapter 4 (paper 3) is a manuscript submitted to the ASCE Journal of Management in Engineering. This paper determines the range of weights for cost that should be used to reach a proper balance between cost and non-cost factors in best-value selection. This research constitutes a unique contribution to both scholarly literature and current practice because it presents a comprehensive analysis of 347 transportation projects using best-value procurement in the United States. The results of this research will facilitate highway agencies the selection of weights, which, in turn, will help to minimize the existing bias toward price of best-value procurement.

Chapter 5 compiles a research summary, details the research contributions and limitations, and exposes future lines of research and researcher reflections.

CHAPTER 2: IMPROVING COST PERFORMANCE IN DESIGN-BID-BUILD ROAD PROJECTS BY MAPPING THE REASONS FOR COST OVERRUNS INTO THE PROJECT PHASES

2.1 INTRODUCTION AND DESCRIPTION OF THE PROBLEM

One of the major challenges facing public administrators over the last 50 years has being how to meet cost goals in public infrastructure projects (Flyvbjerg et al. 2002).

Public administrators use contracting strategies to develop road projects. Through these strategies, they undertake all processes to obtain a complete project design and thereby complete the construction (Touran et al. 2008). Generally, the contracting strategy comprises three elements: project delivery systems, procurement methods, and contract approaches (Molenaar et al. 2014).

In this research, the project delivery system is defined as the model for the processes of designing and constructing infrastructures (Forbes and Ahmed 2011). Traditionally, public administrators have adopted the Design-Bid-Build (DBB) system, which involves contracting the design and the construction independently. The advantages associated with this method are its long history of acceptance and its ability to provide open competition, to establish a clear distinction between roles, to permit flexibility for the owner and to facilitate the process of bidding (Pakkala 2002). However, the separation of the designer and the constructor activities generates some disadvantages, such as the lack of innovation, the linearity of the process, the lack of timely and qualified input from the contractor, and the higher odds of overruns and disputes between parties (AIA, 2007; Brierley, Corkum, & Hatem, 2010; Hasanzadeh, Esmaeili, Nasrollahi, Gad, & Gransberg, 2018; Pakkala, 2002). Moreover, in this project delivery system, public administrators assume most of the risks and are responsible for the errors and omissions of the design (Pakkala 2002).

Some of the flaws inherent in DBB system have been addressed using other methods, such as Design-Build (DB) (AIA 2007; Forbes and Ahmed 2011). This approach implies that the public administrator procures both the design and construction from a single firm (Miller 2000). This method allows more flexibility in developing the design and in procuring the construction of the project (Cheung et al. 2013; Sullivan et al. 2017). It also promotes and enhances communication between the teams, which is crucial to improve project performance (Tran et al. 2017b; US Federal Government 2002). Further, in DB, public administrators generally use best-value procurement to select the teams that will develop the work (AASHTO 2018). In best-value (unlike low bid procurement), decision-makers use several criteria in addition to price to award the contract (Scott et al. 2006). Considering not only price in the selection makes it more likely to select the best contractor possible (Akcay and Manisali 2018; Yu and Wang 2012). Choosing the best contractor possible, therefore, can reduce risks in project development (Morledge and Smith 2013). All these reasons have led DB to become a viable option for the construction of road projects despite the cultural and legislative barriers that it has faced in its implementation (DeWitt et al. 2005; FMI 2018b; Sullivan et al. 2017). Still, some countries encounter legal difficulties in DB implementation in road projects (Alarcón et al. 2018; Lam et al. 2006).

Numerous studies have compared the effectiveness of DB and DBB systems (FHWA 2006b, 2017b; Minchin Jr. et al. 2013; Park and Kwak 2016; Shrestha et al. 2007; Tran et al. 2018b). For instance, Sullivan et al. (2017) is a representative example. These authors analyzed 30 studies—published in the last 20 years—that represent 4,623 projects. These studies contained projects' information about the delivery approach and the cost outcomes. The analysis allowed researchers to compare the cost performance between project delivery methods. The results showed that the DB delivery method is more effective in controlling the cost growth (+2.8%) than

DBB (+5.1%). However, these studies did not examine the possible reasons that lead DB to experience fewer cost overruns than DBB.

Previous research has addressed the reasons for cost overruns in construction projects worldwide including the United States (Bordat, McCullouch, Labi, & Sinha, 2004; Shane, Molenaar, Anderson, & Schexnayder, 2009), Australia (Creedy et al. 2010), Asia (PMI 2014; Rosenfeld 2014; World Bank 2014), South Africa (Khabisi et al. 2016), Europe (Verweij et al. 2015) and South America (Tala González 2015). Most of these research found reasons for cost overruns associated with three broad areas: design, project changes, and procurement.

Design. Shane et al. (2009) identified as one of the main reasons for cost overruns "poor estimating," which, according to these authors, led to "errors and omissions from plans and quantities as well as general inadequacies and poor performance in planning and estimating procedures and techniques." Creedy et al. (2010) analyzed data collected from 231 highway projects in Queensland and found as relevant factors leading to cost overruns: insufficient investigations, deficient specifications and design, and lack of constructability. Another evidence supporting errors and omission in design as a reason for cost overruns is the results of a pilot study developed by the World Bank based on roadworks. It suggested that some of the common reasons for delays and cost overruns were poor preliminary studies, inadequate site investigations and incomplete, delayed or inaccurate designs (with error and omissions) (World Bank 2014). Within the public sector, Khabisi et al. (2016) identified as cost overrun factors: errors and omissions in design, inaccurate estimation of time and cost, and inexact quantities. Another evidence emphasizing design aspects as reasons for cost overruns and delays is Tala (2015). His thesis explored construction projects developed by the Ministry of Public Works in Chile, and one of his findings was that almost 70% of the main reasons for construction cost overruns and delays were influenced by problems in previous phases of the project lifecycle (pre-design and design).

Project changes. According to Love et al. (2012), rework during construction as a result of design changes and errors and omission is a significant contributor to overruns in projects. Several authors agree to consider changes as a fundamental reason for cost overruns. Shane et al. (2009) identified as a reason for cost escalation changes in schedule and scope. They considered that changes in schedule could be due to, for example, design changes or budget constraints. Scope changes may include alteration in design and dimensions as well as increases in project elements. Creedy et al. (2010) found that the first principal factor of cost overrun was design and scope change. For highway projects, the Indian Project Management Institute found that two of the main factors having a higher impact on project cost were the scope creep and the design change (PMI 2014). Rosenfeld (2014) developed a cross-sectional survey for ranking the 15 root causes of construction cost overruns. The second one was "too many changes in owners' requirements or definitions." Khabisi et al. (2016) identified as cost overrun factors within the public sector: change orders, change in scope, and design changes. Verweij and Meerkerk (2015) analyzed a total of 2,804 contract changes in 45 projects (with 84,4% of the projects being roads). They found that scope changes were the most significant reason for cost overruns in transportation infrastructure projects.

Procurement. Several researchers also identified aspects related to procurement. Bordat et al. (2004) found that influential factors of cost overrun of Indiana Department of Transportation (DOT) highway contracts included: the contract bid amount, difference between the winning bid and second bid, and the difference between the winning bid and the engineer's estimate. Shane et al. (2009) mentioned the delivery and procurement approach as a reason for cost overruns.

Rosenfeld (2014) found as the most critical reason for cost overruns having premature tender documents (drawings, bill of quantities, specifications, contracts, and legal documents); Khabisi et al. (2016) identified the policy of accepting the lowest tender in the procurement as one of the main cost overruns triggers within the public sector.

Generally, all these studies resulted in lists and ranking of reasons, some of them suggesting generic recommendations to mitigate those reasons. These studies, however, did not provide a structured analysis that served as a basis to propose those recommendations.

The purpose of this research is to propose a structured approach to provide measures based on DB practices that can help DBB road projects to mitigate the principal reasons that lead to cost overruns. The study contributes to exploring the reasons for cost overruns and their link with some aspects of the project delivery system by identifying and mapping them in the phases of design, procurement, and construction of DBB road projects. This mapping constitutes the point of departure to find specific elements of DB project delivery that might help to minimize these reasons.

An overview of relevant literature has been first introduced. The following section includes the data collection and methods used to analyze the data. Finally, the result, the discussion, and conclusion sections are developed.

2.2 Methodology

To find the main reasons of cost overruns in three phases of a road project—that is, design, procurement, and construction—an exploratory content analysis was performed based on the interviews conducted with 41 professionals involved in road project management in Chile. This study conducted reviewed literature and analysis of DB procurement documents—including DB

Manuals and Request for Proposals (RFPs)—to analyze the identified reasons under the framework of the DB practice in the United States.

This research focuses on the DB delivery system because public administrators have successfully implemented it on delivering road projects. In several cases, DB system saved construction costs, improved quality, and provided benefits associated with innovative solutions (DBIA, 2014).

2.2.1 CONTENT ANALYSIS

According to Krippendorff (1980), content analysis is a research method that allows obtaining replicable and valid inferences from data within the context from which they were taken. The method aims to provide knowledge, new insight, facts illustration, and constructive drivers to action. Fellows and Liu (2015) considered this technique within the domain of construction management. They mention that inductive content analysis may be used to determine the most important aspects of a data set by counting the number of times that a topic is mentioned in that set. Several authors have used content analysis to address construction issues, such as DB evaluation criteria (Xia, Chan, Zuo, & Molenaar, 2011; Xia, Skitmore, & Zuo, 2011), construction contracts (Stanford, Molenaar, & Sheeran, 2016) or performance measures (Harper, Molenaar, Anderson, & Schexnayder, 2014).

This study used the steps for developing inductive content analysis introduced by Elo and Kyngas (2008). Moving through the phases of open coding, creating categories and abstraction, the study was able to determine what the interviewed professionals believed to be the primary causes of cost overruns taking place in each of the design, the procurement and the construction phases of road projects.

2.2.2 Data Collection

This study was part of a major research project awarded in 2017 by the *Centro de Políticas Públicas de la Pontificia Universidad Católica de Chile*. The *Centro de Políticas Públicas* aims to encourage cross-functional work between scholars regarding public issues of general concern. Chilean Director of Public Works, Former Secretary of Public Works, President of the Public Infrastructure Committee of the Chilean Chamber of Construction, as well as professionals from the Department of Public Works and Roads, were involved in the research project and the presentations and seminars conducted in its framework. They followed the progression of the project and provided feedback and validation to the research team.

As part of the project, multiple semi-structured interviews, including three open-ended questions, were conducted with 41 professionals involved in the management of public interurban road projects in Chile (See appendix A for more detail on interviewees' characteristics). These professionals were selected from public administrations (51%) and private companies (49%), based on their level of experience and involvement in the phases of the design, procurement, and construction of road projects in Chile. The average experience of the professionals was 17 years. They were asked the following questions: Question 1 (Q1), Could you identify, at least, three main reasons associated with the design phase that highly impact construction cost overruns? Question 2 (Q2), Could you identify, at least, three main reasons related to the procurement phase that profoundly impact construction cost overruns? Question 3 (Q3), Could you identify at least three main reasons related to the construction phase that highly influence construction cost overruns? The aim of asking these questions was to identify the main reasons affecting cost overruns that could be associated (as events) with each of the phases. The interviewees provided their written responses electronically.

2.2.3 DATA ANALYSIS

The written answers were recorded in the software dedoose (https://www.dedoose.com/), where one document was created for each answer.

In the first step of content analysis, open coding, the researcher followed a traditional approach adopted in social sciences, allowing the codes to emerge from the data (Creswell, 2009). In this research, the codes emerged from the recorded interviews' answers. Each code represented a reason leading to cost overruns. The analysis identified 118 codes. The responses were re-read, aiming to create higher-level codes that gather similar concepts. Afterward, the researches established 78 final codes and listed them with their definition in the research's codebook. Using the research's codebook, researchers assigned the final codes using two independent coders.

According to Neuendorf (2002), reliability "is the extent to which a measuring procedure yields the same results on repeated trials. The notion relevant to content analysis is that a measure is not valuable if it can be conducted only once or only by one particular person." The author states that, in content analysis, when humans assign the codes, the reliability translates in intercoder reliability. The intercoder reliability in this research was measured using Cohen´s kappa coefficient. Cohen's Kappa coefficient (k) is a statistical measure of the level of agreement between two independent raters that takes into account that the possibility of agreement could occur by chance alone (Salabeddin M., 2012). Precisely, k is calculated using the following equation:

$$k=(Po-Pe)/(1-Pe)$$

Where,

Po: is the proportion of the observed agreement between the two raters

Pe: is the proportion of rater agreement expected by chance alone.

K equal to +1 implies complete agreement, and k equal to 0 indicates a lack of agreement between the raters (Salabeddin M., 2012). The intercoder reliability in this research obtained a Cohen´s kappa of 0,72, which corresponds to a good agreement (Domenic, 1994; Fleiss, 1972; Landis & Koch, 2013).

The second step of the content analysis included the creation of categories to describe the phenomenon, increase its understanding, and generate knowledge (Krippendorff, 1980). Affinity diagrams were used to organize and establish groups of codes (Carnevalli & Miguel, 2008). The affinity diagrams sort data and find underlying relationships linking the resulting groups (Andersen & Fagerhaug, 2006).

Each code represented a reason for cost overrun. Researchers ordered the codes by the percentage of agreement between the interviewees. The percentage of agreement was calculated by dividing the number of each code appearance—in each question—by the number of interviewees. Only the reasons with more than 10% of the agreement were considered for the analysis. Questions Q1, Q2, and Q3 focused on each of the design, procurement, and construction phases. However, some of the interviewees' answers referred to reasons for cost overruns associated with other phases not mentioned in the question.

For this reason, the researchers analyzed the group of codes associated with the questions Q1, Q2, and Q3. In each group, researchers filtered each reason by asking where it usually occurred (i.e., whether the reason happened in the design, the procurement, or, in the construction phase). The filtering process resulted in re-assigning codes to different phases from the one that generated the code. For example, question Q1, which referred to reasons in the design phases, had as a codified reason "utility relocation issues." This reason does not occur in the design phase; thus, it was re-assigned to the construction phase. Thus, each code—representing a reason—was associated with the phase in which that code (as an event) occurred.

Finally, the third step of the content analysis was abstraction. This meant to formulate a general description of the research topic through the categories that had been generated in prior steps of content analysis (Robson, 1993). This research builds a discussion upon how the reasons identified in each of the design, procurement, and construction phases can be mitigated by implementing elements from the DB project delivery approach (See Appendix B for more detail on the content analysis process).

2.3 RESULTS

From each of the phases, the researchers obtained three groups of codes representing the main important reasons for cost overruns identified by the interviewees.

When asking for the reasons for cost overruns associated with the design phase (Q1), "Errors & Omissions in the design" was the cause with the highest percentage of agreement—with approximately 40% of consensus among the interviewees. "Inadequate designer capabilities" stood out with roughly 35% of the agreement. The interviewees highlighted specific features associated with the lack of capability, such as "lack of designers' experience" and "lack of designers' local knowledge." Lastly, "Lack of proper basic engineering," achieved 27% of agreement among the participants, this reason includes "poor topography," "poor soils identification and/or geotechnical studies," and "poor general baseline of the area affected by the project."

When asking for the reasons for cost overruns associated with the procurement phase (Q2), the most frequently mentioned cause identified among participants was "vague, incoherent, and with errors Request for Proposal (RFP)" with around 23% of agreement. "Short time to study and clarify the RFP" achieved 13% while "lack of proper information exchange" reached 10% of agreement.

When asking for the reasons for cost overruns associated with the construction phase (Q3), "inadequate constructor capabilities & interest" stood out with roughly 30%. The following main reasons, with 13% of agreement" was related to "project changes." The interviewees specified some of the reasons leading to "project changes," such as "project outdated," "in situ adjustments to real conditions," "change of soil foundation conditions," and "need to include new items during construction." Finally, "long time between design and construction" was mentioned with a 10% agreement among the interviewees.

Figure 3 maps the main reasons identified through the phases of the design, procurement, and construction.

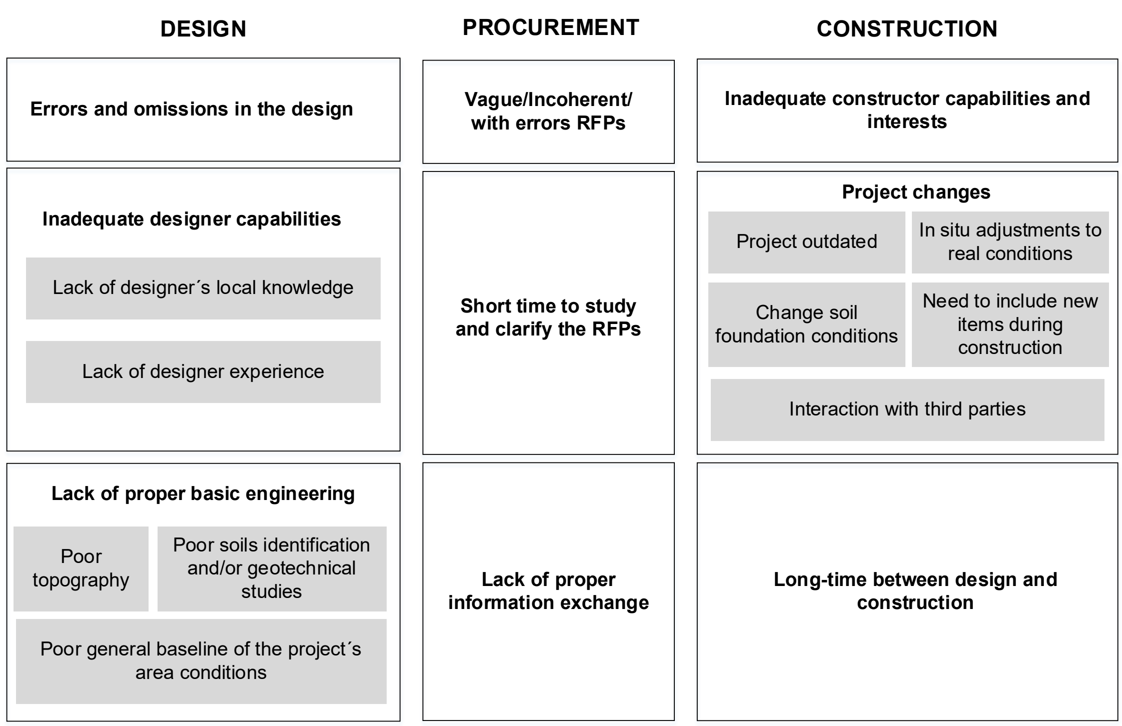


Figure 3 Reasons for cost overruns in design-bid-build road projects.

2.4 Discussion

To discuss how elements associated with DB can be considered to minimize cost overruns in road projects delivered using DBB, the researchers mapped the main reasons for cost overruns in DBB road projects into the design, the procurement, and the construction phases. The reasons for cost overruns obtained in each phase were analyzed under the DB approach to finding elements that could be implemented in that phase of the DBB process to minimize these reasons for cost overruns.

2.4.1 Design Related Reasons

Which DB's practices can be considered in the design phase of DBB road projects to minimize the main reasons for cost overruns associated with this phase? The research identified three main reasons for cost overruns related to the design phase: "errors and omissions in the design," "inadequate designer capabilities," and "lack of proper basic engineering principles." These reasons align with finding from previous research that focused on the reasons for cost overruns in the design (Creedy et al., 2010b; Khabisi et al., 2016; Shane et al., 2009; Tala González, 2015; World Bank, 2014). In the following paragraphs, the authors discuss the design-related reasons under the lens of DB practice.

Errors and omissions in the design. AIA (2007) acknowledges that part of the inefficiencies associated with DBB are related to the separations of responsibilities that lead the design team and the construction team to have different objectives and interests. "Errors or omissions in the design" might be reduced by involving the design and the construction parties in the early stages of the project. Highways England is an example of a road administrator that recognizing relevant problems associated with DBB—including cost overruns around 35%—decided to transition to alternative project delivery methods. They established a long-term plan to introduce changes in the management and delivery of transportation projects. Further, they introduced a variety of strategies, including design-build contracts and early contractor involvement (DeWitt et al., 2005).

Inadequate designer capabilities. The decision-making process needed to select the ideal contractor possible—to achieve the project objectives in the design—requires including additional evaluation criteria besides the price (Akcay & Manisali, 2018). The reasons for having "inadequate designer capabilities" might be minimized by using a selection process that enables road administrators to align the selection of the firms with the project goals. Best-value procurement is a method—currently used in DB road projects—that have the potential of achieving this alignment. In the procurement related section, the details for the application of this method will be addressed.

Lack of proper basic engineering principles and specifically, the "lack of appropriate subsurface investigations." In this regard, DBB entails two shortcomings. First, the lack of ability to obtain timely and qualified contractor's input during the design. These inputs could be related to constructability and scope definition. Second, the designer's lack of knowledge of construction means and methods with the consequent inability to anticipate how contractor choices will influence the behavior of site conditions (Brierley et al., 2010). The DB project delivery with the integration of designer and constructor can provide an excellent context to achieve a fairer contractual understanding in risk allocation for subsurface conditions (Brierley et al., 2010).

In summary, the early integration of the constructor's expertise is the answer to the question stated at the beginning of this section. Road administrators procuring DBB road projects should consider to include the constructor's expertise in the design phase. The contractor's early involvement might help them 1) to reduce errors and omissions in the design as well as (2) to adequately allocate risks associated with gathering adequate information related to basic engineering.

2.4.2 Procurement Related Reasons

Which DB's practices be considered in the procurement phase of DBB road projects to minimize the main reasons for cost overruns associated with this phase? The research identified three main reasons for cost overruns related to the procurement phase: "vague, incoherent and with errors Request for Proposal (RFP)," short time to study and clarify the RFP," and "lack of proper information exchange." Some of these reasons align with finding from previous research that focused on the reasons for cost overruns in procurement (Bordat et al., 2004; Khabisi et al., 2016; Rosenfeld, 2014; Shane et al., 2009). In the following paragraphs, the authors discuss the procurement related reasons under the lens of DB practice.

Vague incoherent and with errors RFP. In the RFPs, road administrators establish the technical and legal requirements that the firms developing the work must accomplish. This document lays the foundations for work performance once the contract is awarded. Thus, clarity and lack of ambiguity are desired to minimize problems in construction. According to Tran et al. (2017), communication between the road administrator and the proposers during the procurement is critical to the success of the projects. In DB procurement in the United States, early exchange of information between the road administrators and the proposers is encouraged. One example of this early exchange of information is the issuance of a draft of RFP for industry review (U.S. Federal Government, 2002). The review process results in the final RFP, in which ambiguities and doubts should have been resolved between the parties involved.

*Short time to study and clarify the RFP and lack of proper information excha*nge. Regarding the former, interviewees explained that the time available once all the information needed is release, is not enough to prepare the proposal. Thus, they set forth that the information exchange process in the procurement is not efficient. In the procurement of DB road projects in the United States, several instances for information exchange between road administrators and proposers are established. One example is informational meetings to introduce general information to the proposers. Another example is the one-on-one meetings to clarify any ambiguous information of the RFP draft. Or, open forum meetings or mandatory pre-bid meetings to summarize and highlight the most relevant aspects of the RFP.

DBB road administrators should tailor—for the procurement process—information exchange activities that fit in their regulatory framework. These activities should aim to facilitate clear and unambiguous information to the proposers so they can prepare proposals adjusted to the road administrator requirements.

In addition to the reasons found in the procurement phase, both in the design and in the construction phases, interviewees identified the "lack of adequate designer and constructor capabilities" as leading causes to cost overruns. A first step towards improving the quality of design and construction teams is to strengthen the selection process that road administrators use to select those teams. As mentioned before, best-value procurement might be an alternative to align the selection process with the project goals. The best-value evaluation takes into account parameters such as time, quality, innovation, and team's qualification and performance. Aspects that require, in turn, the evaluation of the past performance of the team members and the company, the assessment of key personnel's qualifications, and the analysis of evidence related to companies' quality assurance and quality control program (Scott et al., 2006). Assessing, quantifying, and comparing this information enable road administrators to perform a comprehensive evaluation of all the proposers. Therefore, the team that provides better capabilities to the project will stand up among the others, and its selection will contribute to accomplish project goals. According to Morledge and Smith (2013), “selecting the right project team is likely to be the single activity that can reduce most client—road administrator—risks.” They indicate that the best team will promote collaboration in the client’s interest to guarantee the delivery of the project on-time and on-schedule (Morledge & Smith, 2013). Thus, it will be worthy for DBB road administrators to analyze if their current procurement method enables them to select the best team possible to develop the work.

In summary, establishing instances for effective information exchange in the procurement and well as considering goal-based selection methods—in contrast to cost-driven selection methods—is the answer to the question stated at the beginning of this section. Road administrators procuring DBB road projects should consider defining instances for effective information exchange between the road administrator and the proposers. These instances will help to enhance the understanding of the project scope. A better understanding of the RFP will reduce ambiguities and doubts that otherwise would be translated to project changes and cost overruns in the construction phase.

Also, road administrators procuring DBB road projects should consider the use of best-value procurement for selecting the design team and the construction team since this method has the potential of aligning the selection process with project objectives. This alignment would support the selection of the best team to accomplish the road administrator objectives for the design phase and the construction phase.

2.4.3 Construction Related Reasons

Which DB practices can be considered in the construction phase of DBB road projects to minimize the main reasons for cost overruns associated with this phase? The research identified three main reasons for cost overruns related to the construction phase: “inadequate constructor capabilities and interests,” “project changes,” and “long time between design and construction.” Previous research found “inadequate constructor capabilities” as well as “long time between design and construction” reasons associated with the traditional contracting strategy (i.e., DBB and low bid procurement) (Park & Kwak, 2016; Yu & Wang, 2012). Project changes have been profusely identified in previous research on the reasons leading to cost overruns (Creedy et al., 2010; Khabisi et al., 2016; Love et al., 2012; PMI, 2014; Rosenfeld, 2014; Shane et al., 2009; S Verweij & Meerkerk, 2015). In the following paragraphs, the authors discuss the construction-related reasons under the lens of DB practice.

Inadequate constructor capabilities. Similarly to the designer selection, the decision-making process needed to select the best team possible—to achieve the project objectives in construction—requires to include additional evaluation criteria besides the price (Akcay & Manisali, 2018).

The reasons for having “inadequate constructor capabilities and interests” might be minimized by using a selection process that enables road administrators to align the selection with the project goals. Best-value procurement is a method—currently used in DB road projects—that have the potential of achieving this alignment. In the procurement related section, the details for the application of this method have been addressed.

Project changes. Different reasons lead to project changes in construction. Interviewees in this research mentioned causes such as “project outdated,” “in situ adjustments to real conditions,” “changes in soil foundation conditions,” “need to include new items during construction,” or “interaction with third parties.” FHWA (2017) highlighted that changes due to quantity modifications and errors and omissions were higher in DBB projects (+3% cost growth) as compared with similar DB projects (+0.7% cost growth). One of the reasons for the difference in number and magnitude of change orders between these project delivery systems can be the early involvement of builders (Hasanzadeh, Esmaeili, Nasrollahi, Gad, & Gransberg, 2018). Besides, Hasanzadeh et al. (2018) also suggest that the use of best-value procurement would increase the odds of having lower claim costs and higher overall success.

Therefore, incorporating the constructor's expertise in the design phase as well as using the best-value procurement to select the constructor firm might be two measures to include in DBB delivery system to minimize change orders in the construction phase.

A long time between design and construction would be eliminated by using DB system since the designer and the constructor are hired together (Miller 2000). In DBB system, road administrators might consider this aspect by looking for options to create high-level accountability for the design and the construction phases together. This accountability would track and control the duration of the phases as well as the time between them. Further, it would warranty the necessary funding to keep a smoothie progression between the stages of the project.

In summary, early integration of constructor's expertise and establishing high-level accountability of the design, procurement, and construction phases of a project are the answer to the question stated at the beginning of this section. Road administrators procuring DBB road projects should consider to include the constructor's expertise in the design phases. Early contractor's involvement might help them to reduce change orders due to design errors, omissions, quantity variation, and lack of constructability. Further, they should establish high-level accountability that considers (1) controlling the duration of the period between the design and the construction and (2) warranting the funding needed for developing the design and the construction.

Both accountability elements will help road administrators to keep a smoothie progression of the project phases that will attenuate the segregation among them.

2.5 Conclusions and Recommendations

Road projects have a challenge in addressing the causes of cost overruns. The purpose of this research was to help DBB road projects to minimize the main reasons for cost overruns in the design, the procurement, and the construction phases; and, therefore, improve the cost performance of this type of project.

In the design, the main reasons found were “errors and omissions in the design,” “inadequate designer capabilities,” and “lack of proper basic engineering.” In the procurement, these were “vague, incoherent and with errors Request for Proposal (RFP),” “short time to study and clarify the RFP,” and “lack of proper information exchange.” In the construction, the main reasons found were “inadequate constructor capabilities and interests,” “project changes,” and “long time between design and construction.”

The analysis of these reasons under the collaborative approach of DB delivery system suggested four measures that road administrators might implement to minimize the reasons for cost overruns.

1. In the design phase, early integration of the constructor’s expertise. This integration might help (i) to reduce errors and omissions in the design, (ii) to adequately allocate risks associated with gathering adequate information related to basic engineering, and (iii) to minimize change orders due to design errors, omissions, quantity variation, and lack of constructability.

2. In the procurement phase, there is a need for effective information exchange between the road administrator and the proposers. These instances might help to enhance the understanding

of the project scope. At this stage, a better understanding of the RFP will reduce ambiguities and doubts that otherwise would be translated to project changes and cost overruns in the construction phase. As Cheung et al. (2013) stated: “the improvement of information flow would likely improve project performance.”

3. In the procurement phase, there is a need to use a best-value approach to select the design team and the construction team. This procurement method considers different evaluation criteria besides prices and thereby enables road administrators to align the selection process with the project objectives. This alignment drives the selection of the best team, which will promote collaboration in the client’s interest to guarantee the delivery of the project on-time and on-schedule (Morledge & Smith, 2013).

4. In general, establishing one point of high-level accountability for the design, procurement, and construction phases of a project might help these phases to keep a smoothie progression that, in turn, would attenuate the segregation among them.

This high-level accountability would comprise (1) controlling the duration of the period between the design and the construction and (2) warranting the funding needed for developing the design and the construction.

This research contributes to the knowledge of construction management by building upon previous research on the reasons for cost overruns. Previous research derived lists of reasons for cost overruns and generic mitigation measures (Bordat et al., 2004; Creedy et al., 2010; Khabisi et al., 2016; KPMG, 2014; Kumaraswamy, 2014; Rosenfeld, 2014; Shane, Molenaar, Anderson, & Schexnayder, 2009; Tala González, 2015; Stefan Verweij et al., 2015). However, much of this work did not provide a structured analysis as a basis to formulate those measures. This research identified the main reasons for cost overruns in road projects. The study used the mapping of these reasons into the phases to provide a framework to discuss mitigation measures through the lens of DB practice.

This study will serve road administrators to start the transition from DBB delivery system to more collaborative approaches that will help minimize the reasons for cost overruns as well as other DBB challenges related to quality issues, inspections' reinforcement, guarantees, and responsibilities. Reducing these aspects, will, in turn, improve the overall project performance.

The main limitation of this research relates to the procedures to collect a portion of the data for this study. The authors used semi-structured interviews. This procedure inhibited researchers' ability to conduct a more throughout validation of the results. This limitation had a slight impact because the authors validated the results by triangulating the findings with the ones obtained from international research. However, future research should be further developed in the identification of cost overruns using case studies that deepen into technical aspects evaluation as well as enable a follow up of the projects analyzed.

Also, this study suggests future research to explore and compare the links existing between the reasons for cost overruns and the elements comprising different alternative project delivery systems.

Chapter 3: Structured Approach for Best-Value Evaluation Criteria: U.S. Design-Build Highway Procurement

3.1 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 constitutes best-value. Otherwise, proposers approach the procurement with a low-bid mentality (Tran et al. 2018a). 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.

3.2 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.

3.2.1 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.

3.2.2 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 lead to minimizing 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 4 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.

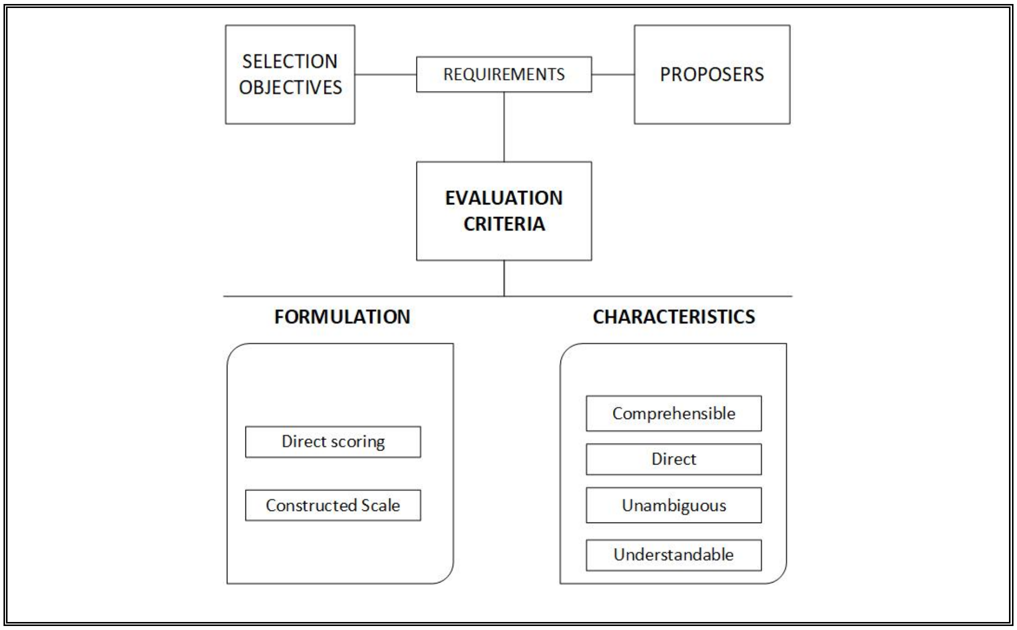


Figure 4 Decision analysis theory concepts to consider in evaluation criteria formulation

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.

3.3 Research Approach

This research applied a structured approach to creating 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) consists 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 changed 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 researchers "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 5 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 (see appendix C for an extensive summary of the RFPs analyzed).

Table 5 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

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 (See appendix D for additional content analysis detail)

3.4 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.

3.4.1 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 of 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.

3.4.2 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 6".

Table 6 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.

(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 7 shows how this evaluation criterion is formulated.

Table 7 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.

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.

3.4.3 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 8 shows:

Table 8 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

"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 9".

Table 9 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

*"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.

3.4.4 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 5).

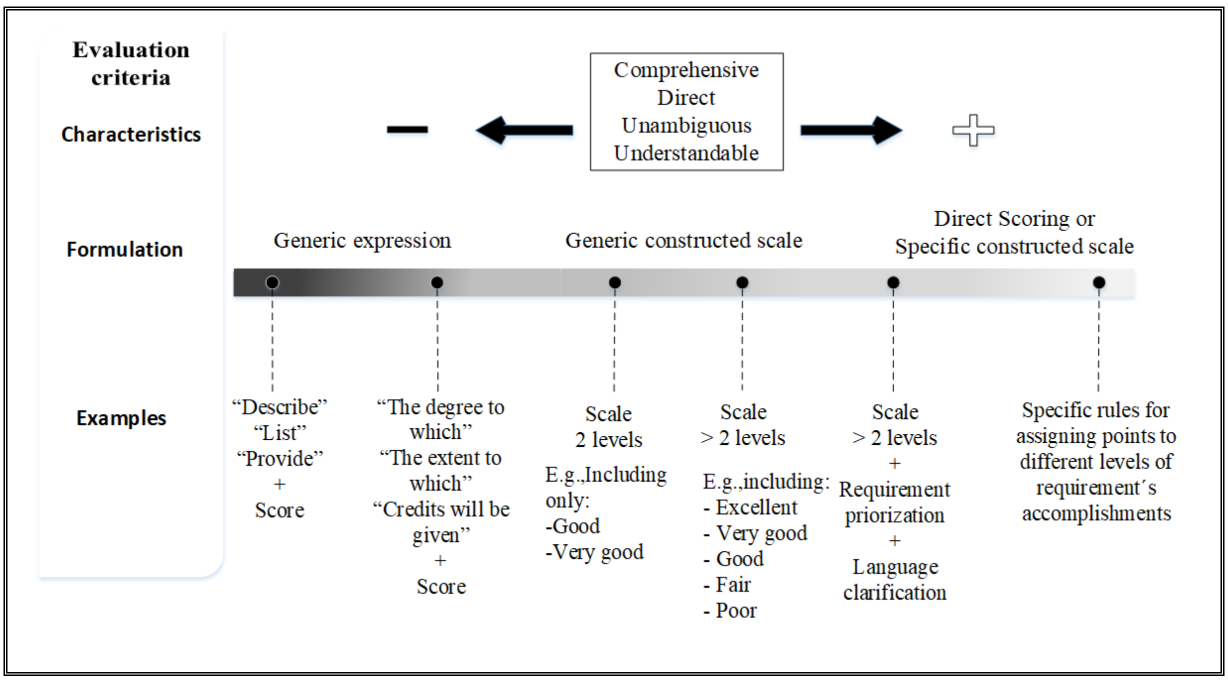
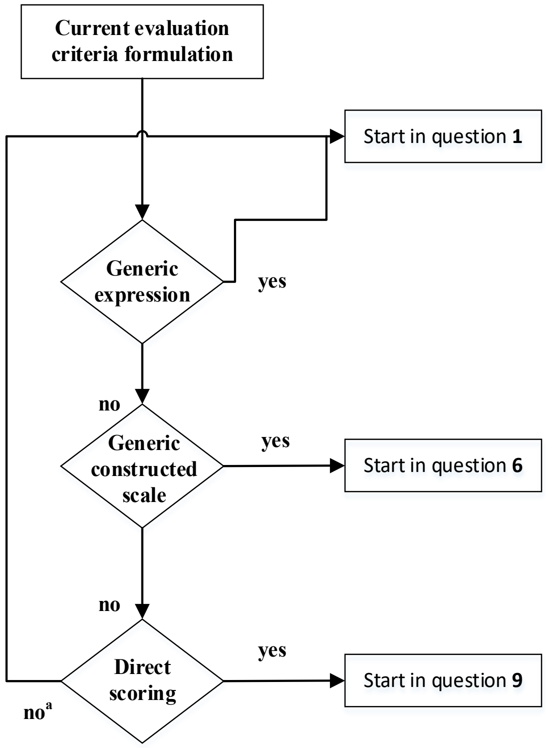


Figure 5 Evaluation criteria spectrum

3.5 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 6). 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.



[a] In case the evaluation criterion do not follow any of the three categories, this research recommends to start in question 1 .

KEY QUESTIONS

1. What do the evaluation criteria measure in the proposal?

2. What documentation is required from the proposers?

3. Does the number or definition of the requirements enable a proper assessment?

4. Are the requirements too many/too general that it is challenging to conduct a detailed evaluation of them?

5. Is it possible to asses the requirements by the direct assignment of scores or by a constructed scale?

6. Is it possible to differentiate the levels of accomplishment of each proposal in the evaluation of the requirements?

7. What are the levels of accomplishment of each requirement?

8. Is it possible to know to what extent a proposal reaches the requirements by recognizing the results of the evaluation in terms of le or score?

9. Can different evaluators reach the same interpretation/measurement?

10. Is the language used to define the rules of score/level's assignment precise?

Figure 6 Proposed structured approach for writing consistent evaluation criteria

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.

3.5.1 Generic Expression Example

Table 10 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 10 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.	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.

• State the degree to which the approach includes final commitments and list any specific commitments in a table.	• Safe conditions for pedestrians and other people accessing the Project Right of Way during construction other than those traveling by vehicle.

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 assess 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 11 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 impacts’ 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 11 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.*

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, and 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.

3.5.2 Generic Constructed Scale Example

Table 12 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 12 Safety. Adapted from KYTC. Boone. Route I-275 Project (2019)

Requirements	Evaluation criterion formulation: Generic constructed scale	
• Describe the safety considerations specific for this project. • Discuss the firm's overall approach to safety	**Scoring range**	**Description**
	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.

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 13. 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 13 Safety. Requirements and evaluation criterion re-formulated

Requirement (Reformulated)	**Evaluation criterion formulation: Specific constructed-scale (Reformulated)**	
• 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.	**Scoring range**	**Description**
	90-100	Proposals in this range identify the most relevant risks providing a comprehensive rationale 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 rationale 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 some of the relevant risks providing a rationale for their probability of occurrence and

o Prevention and mitigation measures. Including associated costs		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 rationale 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

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.

3.5.3 DIRECT SCORING EXAMPLE

Table 14 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 14 Maximize overall safety, capacity, and operation. Adapted from CDOT. Cimarron project (2014)

<table>
<tr><th>Requirements</th><th>Evaluation criterion formulation: Direct scoring</th></tr>
<tr>
<td>Design-builders should submit the Additional Requested Elements (ARE) that they consider among the following proposed by this Agency:

1. Full-Width I-25 Bridges

2A. Widen US 24 Bridge over Fountain Creek and provide Additional lanes to 8th

2B. Replace US 24 Bridge over Fountain Creek and provide Additional lanes to 8th

3. Trail and Creek Improvements along Fountain Creek up to 8th street

4. Contractor Defined ARE (additional operational Improvements on US 24 and at the I-25 and US 24 Interchange</td>
<td>Each ARE has different points associated as Table 7 shows:
<table>
<tr><th>Additional Requested Elements (ARE)</th><th>Points</th></tr>
<tr><td>1. Full-Width I-25 Bridges</td><td>3</td></tr>
<tr><td>2A. Widen US 24 Bridge over Fountain Creek and provide Additional lanes to 8th</td><td>4</td></tr>
<tr><td>2B. Replace US 24 Bridge over Fountain Creek and provide Additional lanes to 8th</td><td>13</td></tr>
<tr><td>3. Trail and Creek Improvements along Fountain Creek up to 8th street</td><td>2</td></tr>
<tr><td>4. Contractor Defined ARE (additional operational Improvements on US 24 and at the I-25 and US 24 Interchange</td><td>7</td></tr>
<tr><td>Maximum Subtotal points (ARE 2A and 2B are Mutually Exclusive)</td><td>25</td></tr>
</table>
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</td>
</tr>
</table>

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 15):

Table 15 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.

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.

3.6 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 on 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.

3.7 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.

CHAPTER 4: THE IMPORTANCE OF NON-COST CRITERIA WEIGHTING IN BEST-VALUE DESIGN-BUILD U.S. HIGHWAY PROJECTS

4.1 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; Gransberg and Ellicott 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 U.S. for the last two decades (Tran et al. 2017a). 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. 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. 2008b); 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 and scores 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.

4.2 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 2018b; 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.

4.3.1 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 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.

4.3.2 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 regards 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.

4.3.3 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. 2017a). 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. 2017a), 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.

4.3 Research Methodology

The research followed a four-step process (Figure 7). 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.

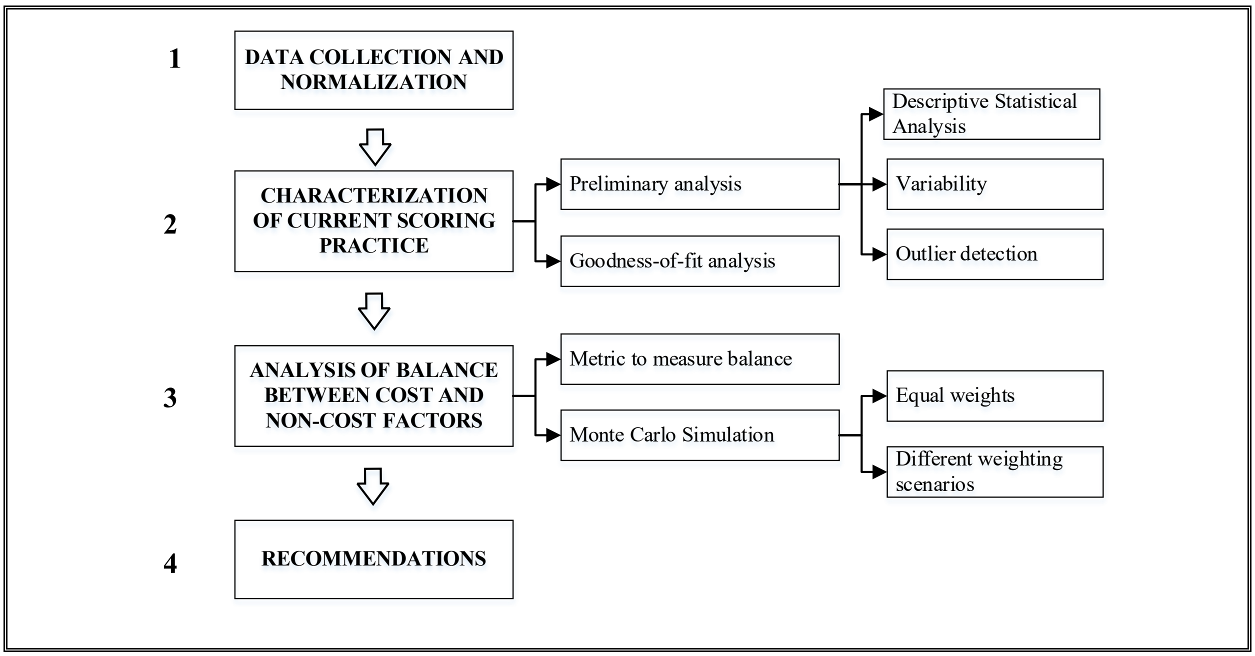


Figure 7 Research methodology

4.3.1 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 requirements 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.

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.

4.3.2 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 analyses (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 issues 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.

4.3.3 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 scenarios of weights. 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 (Figure 8). 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.

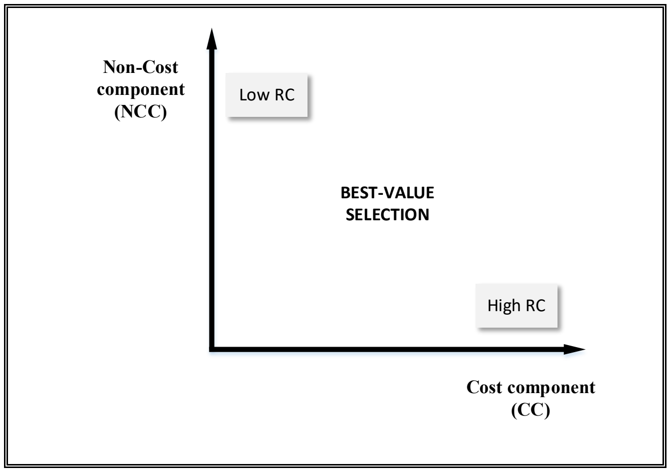


Figure 8 Cost component vs. Non-cost component

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.

Based on current practice, the authors determined 0.3 and 0.7 as the extreme values for the weight of cost. 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.

4.4 Results

4.4.1 Characterization of Current Scoring Practice

Preliminary Analysis

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 16). The data spread is higher in cost scores (standard deviation of 0.103 vs. 0.079 in Table 16).

Table 16 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

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. Figure 9 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).

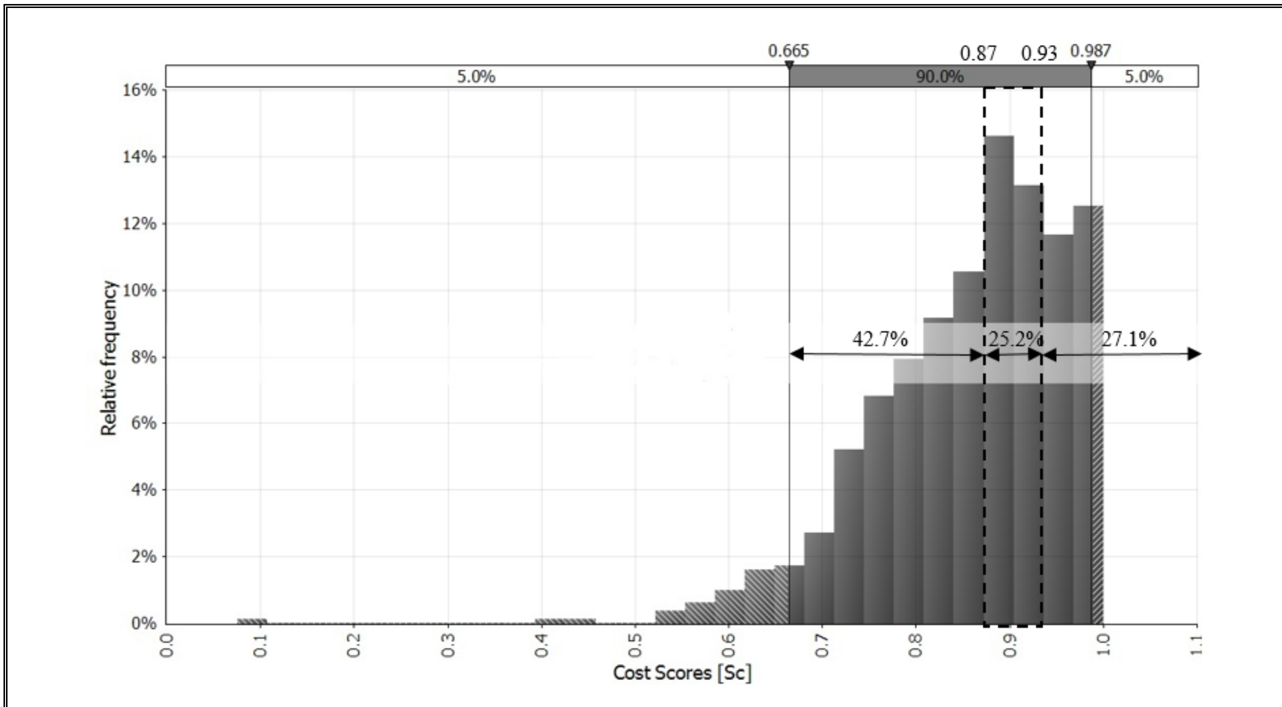


Figure 9 Cost Scores. Relative Frequency

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, Figure 10 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.

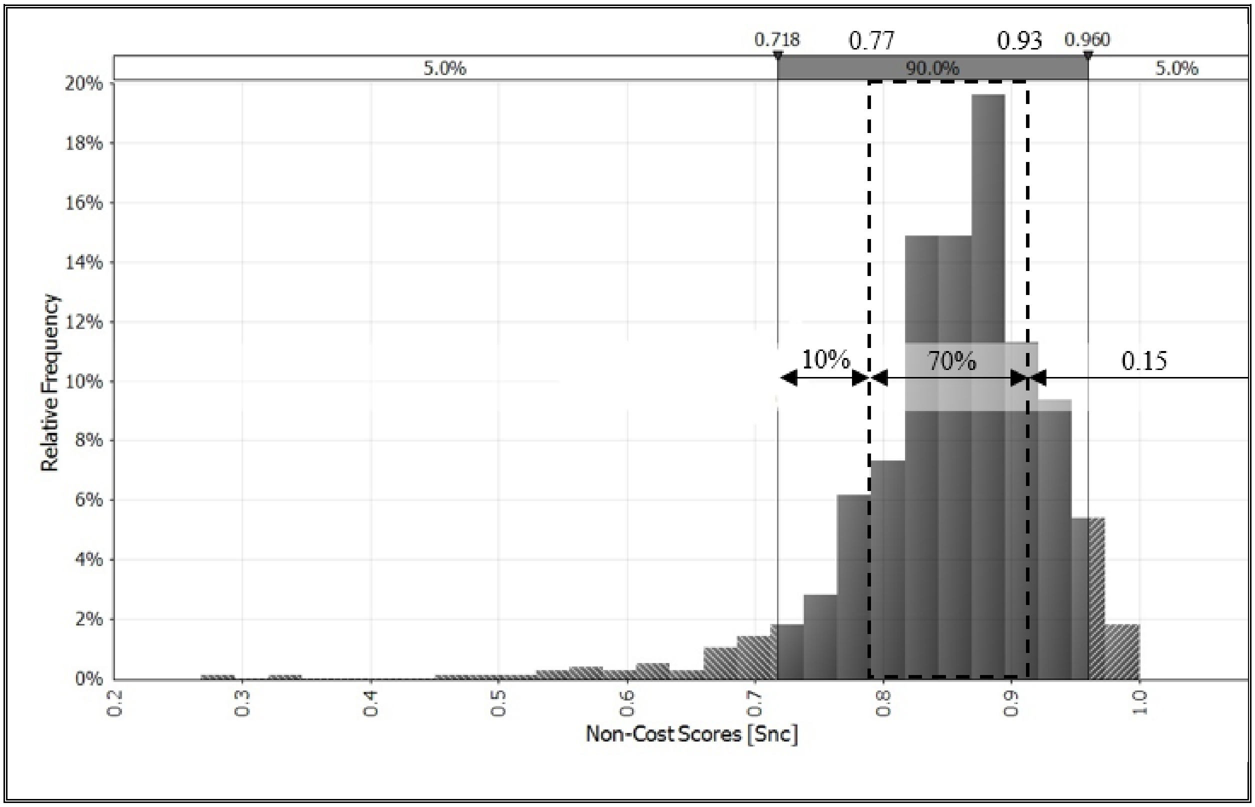


Figure 10 Non-cost scores. Relative Frequency

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 (See Appendix E for more detail on variability and outlier's 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 17).

Table 17 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

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 (See appendix F to check the goodness of fit graphs for both distributions).

4.4.2 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 Figure 11, where RC is higher than “1” in 68.6% of the cases.

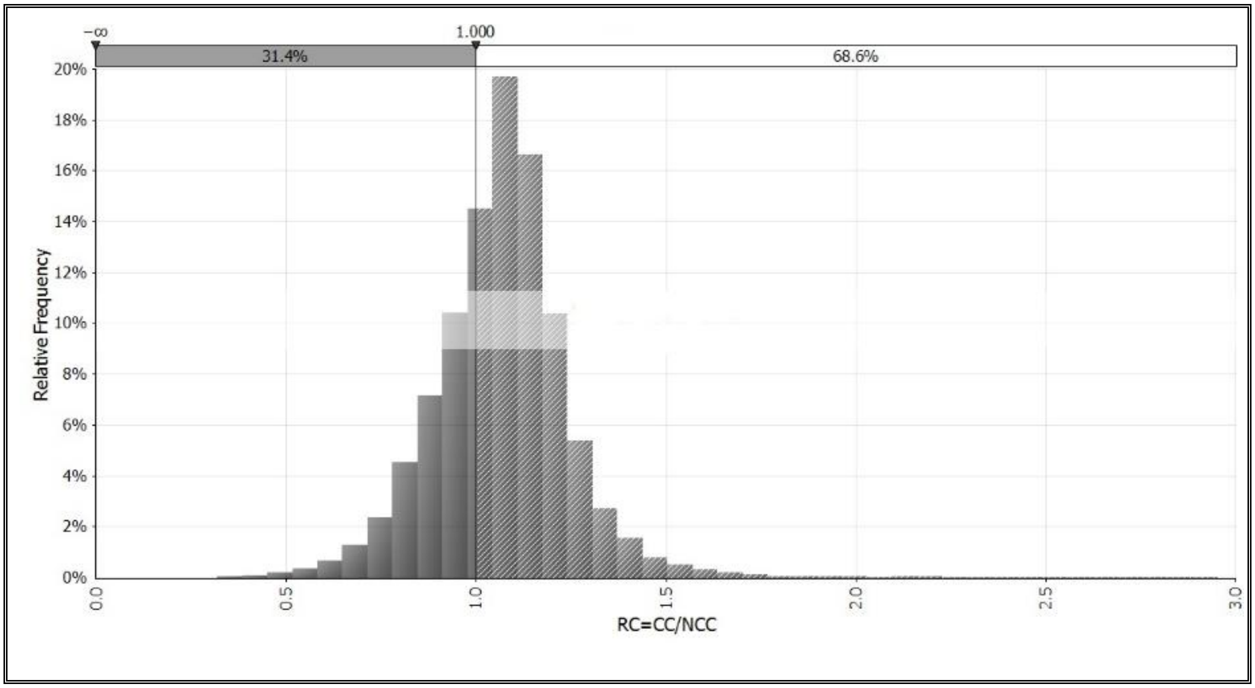


Figure 11 Relative frequency graph when assigned equal weights (Wc = Wnc = 0.5)

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 (Figure 9, Figure 10).

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

Figure 12 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 5th and 95th percentile are recorded. These values, plotted in Figure 12 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.

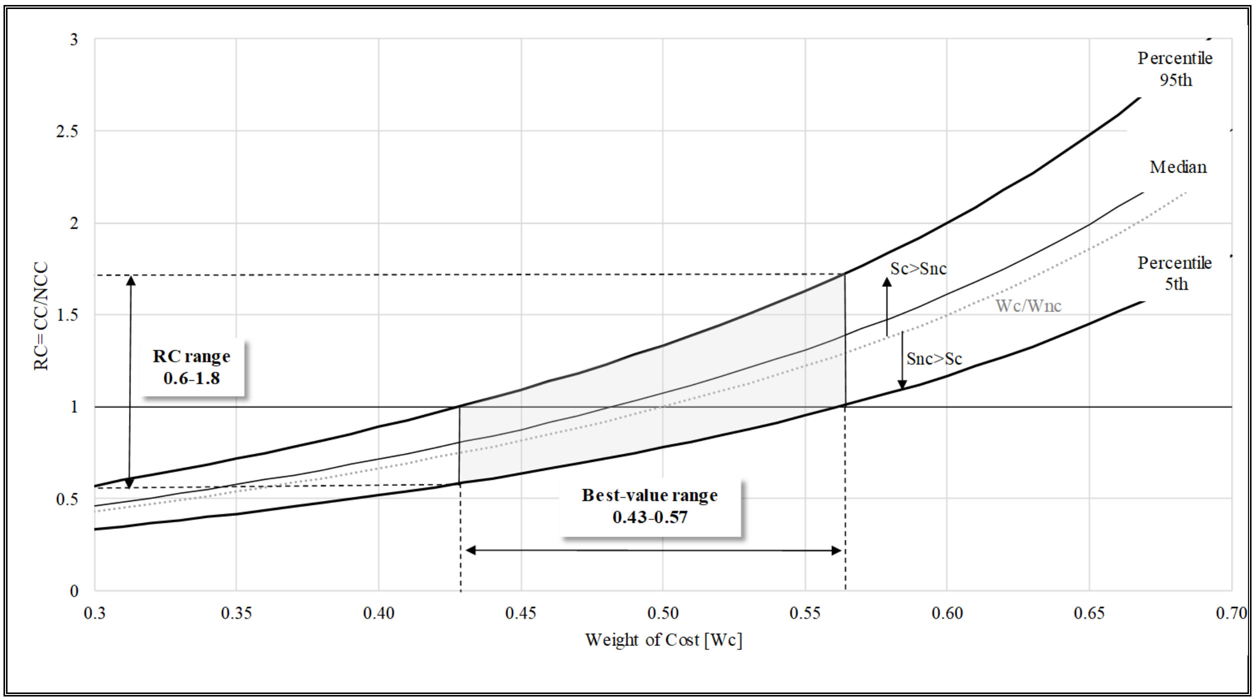


Figure 12 Ratio Cost/Non-cost components [RC] vs. weight of cost [Wc]

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, Figure 12 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 Figure 12. 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.

4.5 Discussion

This research aimed to find the ranges of weights and scores 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.

4.5.1 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 balanced range.

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. 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 proposes 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%.

4.5.2 Scores

As well as weights, the scores might play an essential role in characterizing the best-value selection. As previously said, the proposed best-value range of weight of cost between 0.43 and 0.57 is due to the existing scoring tendencies. The 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.

4.6 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.

CHAPTER 5: CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

5.1 CONCLUSIONS

This thesis started by showing the need to revisit current DBB-low bid and DB-best-value knowledge and practice in highway projects. This need led to posit two intuitions (Figure 13) that have been addressed by answering three research questions.

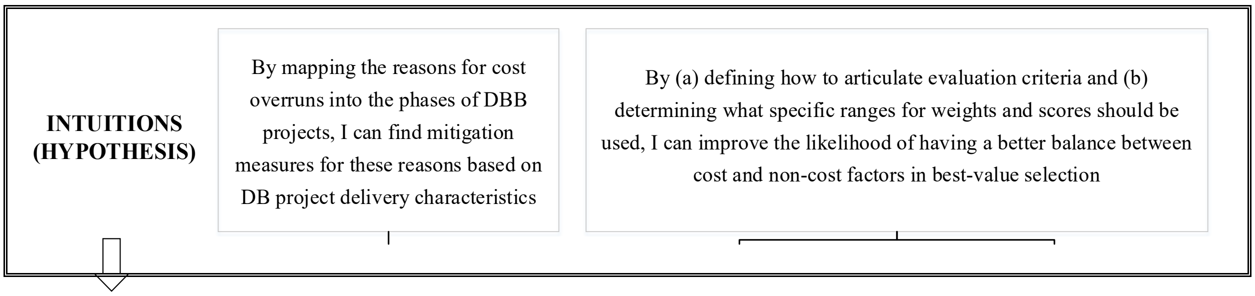


Figure 13 Intuitions (hypothesis)

The first intuition was addressed by answering the first research question. This question aimed to find measures based on DB strategies that served to mitigate the main reasons for cost overruns into the design, the procurement, and the construction phases of DBB projects.

The firs research (Chapters 2) provided a mapping of the primary reasons for cost overruns into the phases of design, procurement, and construction (Figure 14) by examining 41 interviews using inductive content analysis.

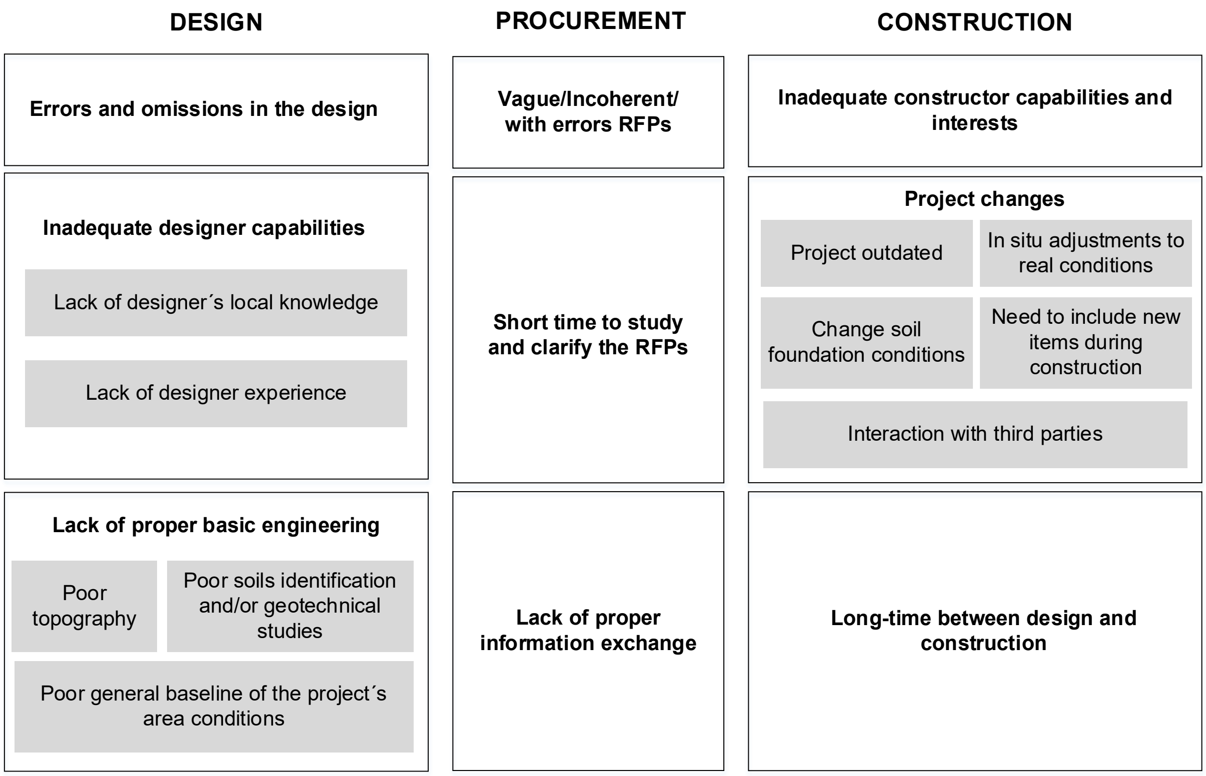


Figure 14 Reasons for cost overruns in design-bid-build road projects.

The study determined four main mitigation measures based on the DB delivery system that might help to mitigate the three main reasons for cost overruns identified in each of the phases (Table 18).

Table 18 Mitigation measures and associated reasons

Mitigation measures	Reasons for cost overruns
# 1. Early contractor involvement *(AIA 2007; Brierley et al. 2010; DeWitt et al. 2005; FHWA 2017b; Hasanzadeh et al. 2018)*	Error & Omissions in the design; Lack of proper basic engineering; project changes
# 2. Best-value selection *(Akcay and Manisali 2018; Morledge and Smith 2013; Scott et al. 2006)*	Inadequate designer and constructor capabilities
# 3. Effective information exchange in the procurement *(Tran et al. 2017b; US Federal Government 2002).*	Vague Request for Proposals (RFP), short time for clarifications, lack of proper information exchange
# 4. High-level point of accountability *(Miller 2000).*	Long time between design and construction

The second intuition was addressed by answering the second and third research questions. These two questions explored evaluation criteria, weights, and scores in best-value procurement in order to find ways in which these elements could help to reach a better balance between cost and non-cost criteria in the best-value selection.

Through answering research question 2 (chapter 3), the research found and structured approach to write consistent, and therefore, more influential evaluation criteria. This structured approach reflects the interactions between decision theory analysis and current practice on evaluation criteria. Depending on where each highway agency's evaluation criteria lie 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.

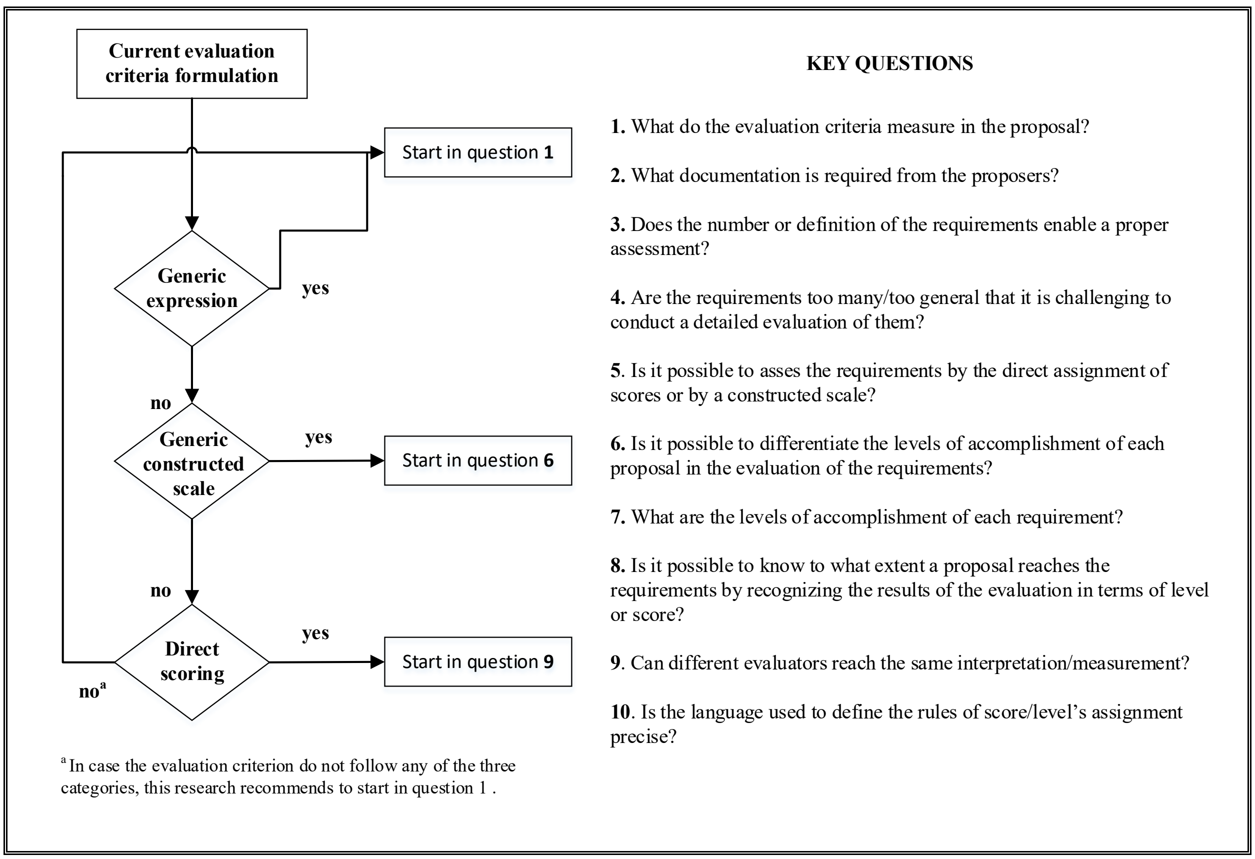


Figure 15 Proposed structured approach to write consistent evaluation criteria.

Finally, through answering research question 3 (chapter 4), the study found that ranges of weight higher than 57% lead to selections that are biased toward cost. Using Monte Carlo simulation and the largest dataset, to date, of best-value scores, this research found a best-value range for the weight of cost between 43% and 57% (Figure 16).

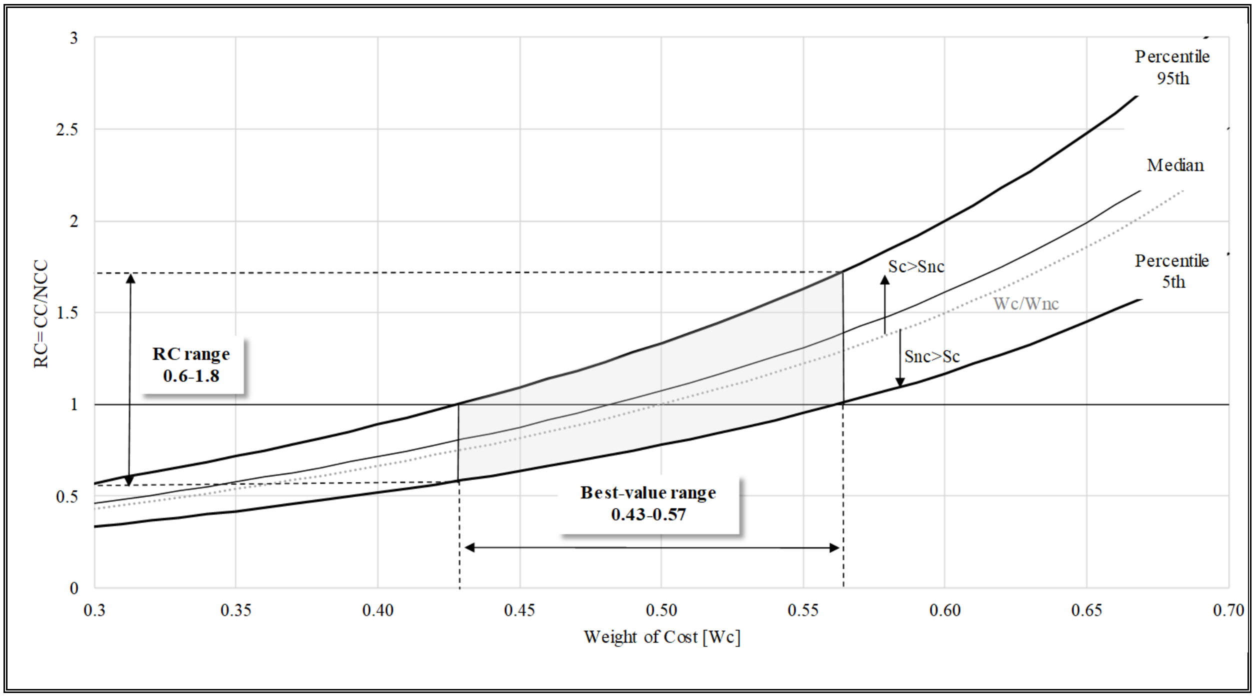


Figure 16 Ratio Cost/Non-cost components [RC] vs. weight of cost [Wc]

In order to balance cost and non-cost criteria, the study also recommended highway agencies to widen the range of non-cost scores and to be aware of the cost dispersion when defining the weights of cost in each procurement.

The answers to questions 2 and 3 served, therefore, to improve the likelihood of having a better balance between cost and non-cost factors in best-value selection.

Figure 17 summarizes this research's intuitions, research questions, and research answers.

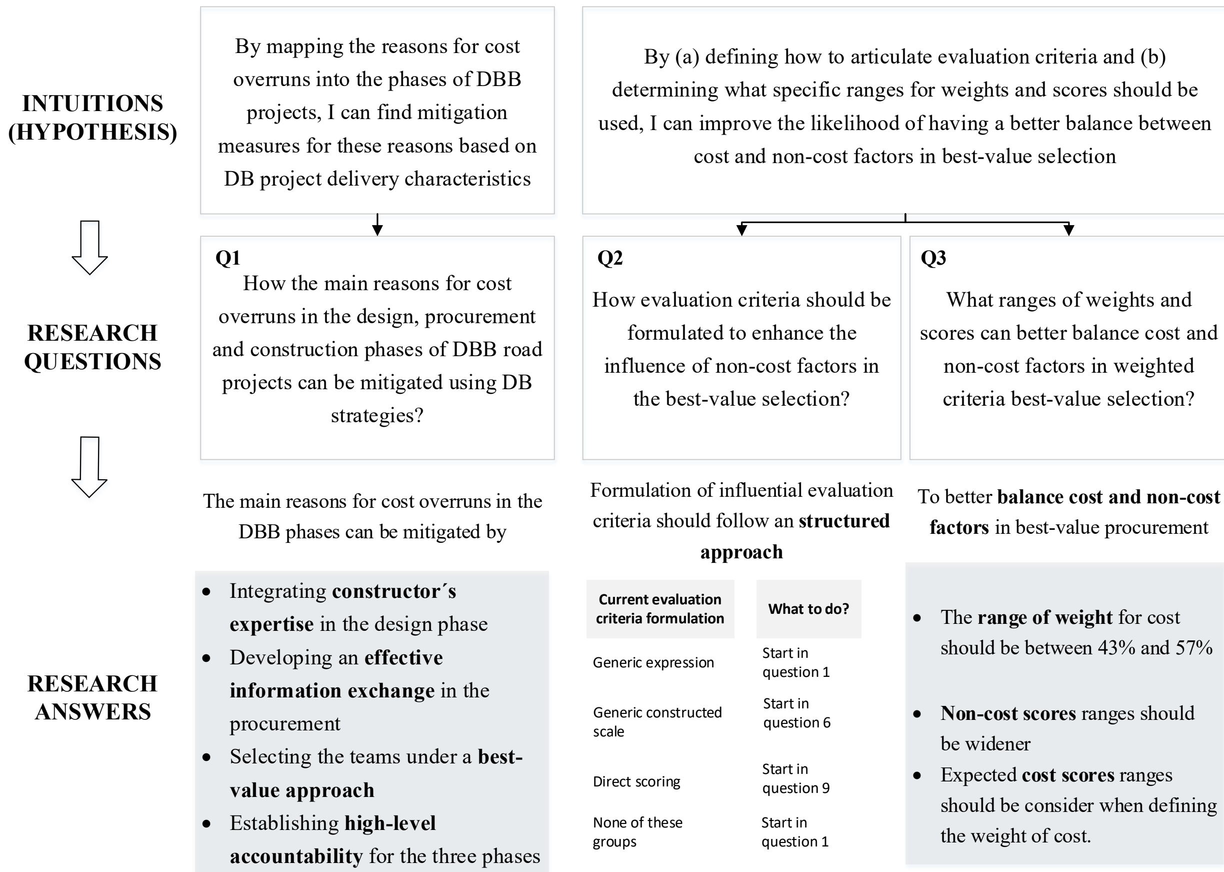


Figure 17 Intuition-Research Questions-Research Answers

5.2 CONTRIBUTIONS

This research contributes to the body of knowledge of cost overruns in design-bid-build road projects and best-value procurement for design-build highway projects. The three independent research developed in this thesis resulted in the theoretical and practical contributions showed in Figure 18.

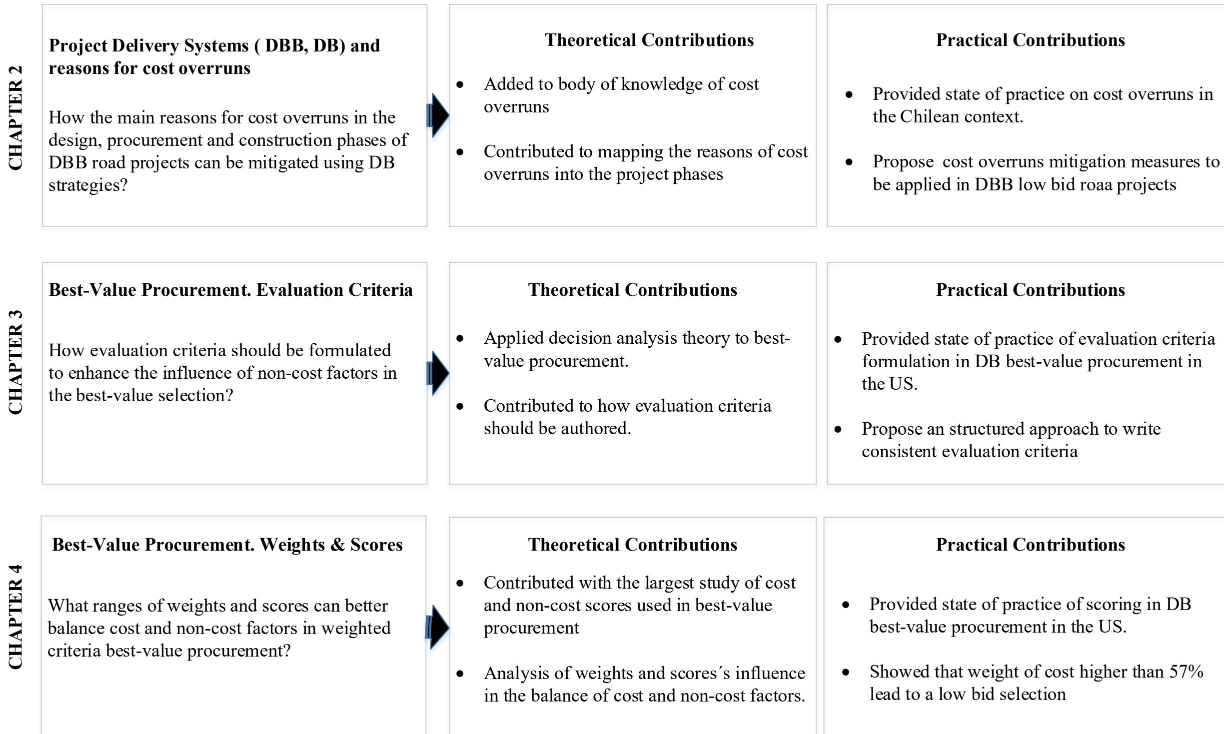


Figure 18 Thesis organization and contributions

This research also is a unique contribution to the improvement of current state-of-the-practice of highway agencies managing the design and the construction of highway projects.

5.2.1 Contributions to the Theory

Chapter 2 contributes to the body of knowledge of cost overruns in construction projects by building upon previous research on the reasons for cost overruns and mapping them into the project phases. Previous research derived lists of reasons for cost overruns and generic mitigation measures (Adam et al. 2017; Al-Hazim et al. 2017; Alarcón et al. 2018; Arditi et al. 1985; Baloi and Price 2003; Bhargava et al. 2010; Creedy et al. 2010a; Khabisi et al. 2016; Rosenfeld 2014; Shane et al. 2009; Sohu et al. 2017; Tala González 2015; World Bank 2014). However, much of this work did not provide a structured analysis as a basis to formulate those measures. For example, Rosenfeld (2014) identified the main reasons for cost overruns globally in the construction industry, and Tala (2015) found the main reasons occurring in Chilean infrastructure projects.

However, none of them considered an structured approach based on their findings to propose specific measures to minimize these reasons. This research expanded the previous studies on cost overruns by not only identifying but mapping the reasons for cost overruns into the project's phases to provide a framework to discuss mitigation measures through the lens of DB project delivery system.

Given that the project delivery systems articulate the management of the project phases, mapping the reasons for cost overruns into these phases facilitates the proposals of measures that can be applied in each of the phases. This approach serves to incorporate measures that encourage collaboration in the phases of DBB projects.

Chapter 3 and 4 contribute to the body of knowledge of alternative contracting methods procurement. Chapter 3 provides a structured approach to authoring best-value evaluation criteria. Previous research on best-value procurement focused on the identification and categorization of evaluation criteria, but not on their formulation (De_Araújo et al. 2017; Molenaar and Johnson 2003b; Molenaar and Tran 2015; Scott et al. 2006; Tran et al. 2017a; Xia et al. 2011b; a). For example, De Araújo et al. (2017) categorize evaluation criteria used in different types of projects. These authors found that for highway projects, the criteria most used were cost, time, quality, staff features, and financial. Xia et al. (2011a), on the other hand, explicitly focused on the type of criteria used in design-build projects. They found that for highway projects, the most relevant criteria were cost, experience, and management. This research expanded the previous studies on evaluation criteria by not focusing on what criteria categories are used (i.e., what is measured) but on how the criteria should be written to be effective in the evaluation (i.e., how is measured). Specifically, this research studies current practice on best-value evaluation criteria formulation.

Based on this analysis, the research contributes by creating a structured approach for writing consistent evaluation criteria grounded on established decision analysis theory.

This new structured approach comprises ten (10) questions that guide the process of writing evaluation criteria by linking project goals, requirements, and evaluation criteria characteristics. The structure enables highway agencies to start with different questions depending upon the pattern of writing category that has also been identified in this research.

Chapter 4 analyzed a unique data set of 882 non-cost scores and 1,148 cost scores from 347 best-value projects aiming to explore how the actual scoring trends influence the balance between cost and non-cost factors in best-value selection. Previous studies focused on the analysis of selection elements such as weights and scores with purposes that differ from the analysis of the balance between cost and non-cost factor in the selection (Abdelrahman et al. 2008b; Asmar et al. 2010; Ballesteros-pérez et al. 2015, 2016; Barron and Barrett 1996; González-Cruz 2012; Lorentziadis 2010; Saaty 1980). Based on the characterization of current score trends and the simulation of the weighted criteria algorithm, this research proposes the range of weight for cost that should be used to reach a proper balance between cost and non-cost factors in best-value selection.

The range of weight for cost is defined between 43% and 57%. In this range, the scores given to cost and non-cost components can be influential in the selection. Further, the cost component is constrained to be between 0.6 and 1.8 the non-cost component. This means that selections in the best-value range are based on balanced criteria. These results are completely new because, until now, there was not enough data to analyze scoring trends in best-value procurement.

Overall, this thesis contributes to providing research-based proposals that aim to improve the current contracting strategies.

5.2.2 Contributions to Practice

Chapter 2 contributes to current practice in DBB road projects by providing four measures to mitigate the main reasons for cost overruns identified in these types of projects. This research will serve road administrators to start the transition from DBB to more collaborative approaches that will help minimize the reasons for cost overruns as well as other DBB challenges related to quality issues, inspections' reinforcement, guarantees, and responsibilities. Chilean public road administration could benefit from this research by considering the proposed measures in the design, procurement, and construction phases of road projects.

Some examples of incorporating the four mitigation measures in current DBB practice in Chile are summarized in Table 19.

Table 19 Proposed mitigation measures. Examples for implementation

Mitigation measures	Examples for incorporating mitigation measures in DBB practice
# 1. Early contractor involvement	-Constructor hired by highway agency in the design phased (bridging see section 5.4) -Required the design team to have an experience constructor professional on the team
# 2. Best-value selection	-in the design phase, it is needed to explore the current best-value effectiveness in order to propose specific examples. -the construction, incorporation of Alternative Technical Concepts (ATC) in the procurement
# 3. Effective information exchange in the procurement.	**-**Consider a qualified team for developing the procurement. This will enable - Organize more meeting for clarifying RFP concepts - Establish an effective industry review of the RFP.
# 4. High-level point of accountability	-Appoint a project leader who oversight the design and construction of one project.

Chapter 3 and 4 contribute to minimizing the existing bias toward cost in best-value procurement by improving the formulation of evaluation criteria as well as by recommending ranges of weights and scores that help balance cost and technical factors in the best-value selection. Specifically, the structured approach proposed in Chapter 3 might help improve current practices in authoring consistent evaluation criteria. 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 be more influential in the selection process. Second, consistent evaluation criteria can precisely show 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.

An example of the application of the structured approach in real evaluation criteria is explained in section 3.5 of this thesis. This section includes examples of evaluation criteria categorized as generic expressions, generic constructed scale, and direct scoring.

The range of weight for cost recommended in Chapter 4 will help improve existing highway agencies' practice by facilitating the selection of weights. Currently, there are no standard guidelines, and each agency decides on the basis of its knowledge and expertise. This situation has shown to lead best-value procurement to a low bid selection, which is precisely the opposite result that is expected from best-value procurement. Further, this research will raise awareness about the relevance of determining adequate weights for cost as well as of the need to think holistically about expected price bids (depending on the type of project) and ranges of non-cost scores and weights when planning any best-value procurement. The US. DOTs could benefit from Chapter 3 and 4 research by adopting the structured approach for writing evaluation criteria as well as considering the suggested weight for cost and recommendations on scores in best-value procurement.

En Chile, this research might also raise awareness about how effective is the best-value procurement currently used to select the design teams.

5.3 LIMITATIONS AND FUTURE RESEARCH

This section summarizes the limitations associated with each of the three research as well as the future research lines that will address these limitations. Further, there are some questions raised during the development of the research, but that remain unanswered. These questions are here considered as points of departure for future research.

In research 1 (Chapter 2), the main limitation relates to the procedures used to collect the data. The study used semi-structured interviews. This procedure inhibited the researcher' ability to conduct a more throughout validation of the results. This limitation had a slight impact because the authors validated the results by triangulating the findings with the ones obtained from international research. However, future research should be developed in the identification of cost overruns using databases (which might gather project characteristics and cost overruns data) and case studies. The analysis of these sources of information will aim to validate the results of this research as well as to find patterns leading to the root causes of cost overruns.

Research 1 proposes mitigation measures that are based on best practices of best-value procurement in the US. However, these measures need to be tested in Chilean contexts. To this end, future research should, first, survey road administrators' wiliness to implement specific measures based on the four ones proposed on this research (the ones included in Table 19 can serve to start). Second, conduct case studies to analyze opportunities, challenges, and effectiveness.

What are the reasons that make Latin American countries be more delayed in the development of policies that include collaborative approaches in the management of the design and construction of highway projects is a question that remains unanswered.

Future research should explore the contextual factors that prevent collaborative approaches to be accepted.

Research 2 (chapter 3) was 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.

The development of this research also left some questions unanswered. For example, how do highway agencies determine project goals in each procurement? How do agencies determine evaluation criteria based on the goals? Does this link really exist? To answer these questions, future research needs to find and test decision-making methods that can serve to select project goals and evaluation criteria.

Another question raised when thinking of different collaborative approaches. Besides DB, other approaches such as Constructor Manager at Risk or Alliance have been used in highway project delivery. Several studies have focused the analysis on comparing cost performance among delivery systems. However, how (in each of these cases) highway agencies select the teams, what evaluation criteria are used, and how these elements might influence a better selection of teams needs to be explored. Thus, future research should compare evaluation criteria across collaborative approaches aiming to find synergies.

Research 3 (chapter 4) focused on the exploration of the balance between cost and non-cost factors in a unique award algorithm, the weighted criteria. Highway agencies procuring best-value projects have used other types of award algorithms, such as adjusted scores or adjusted bid. Different award algorithms might lead to different balance between cost and non-cost factors. Therefore, future research should be developed to explore the influence of the award algorithm on the selection results.

The results of research 2 and research 3 lead to questioning whether different evaluation criteria formulations conduct to different non-cost score ranges and, in turn, to different balance between cost and non-cost factors. Future research should compare scoring trends across different evaluation criteria formulations.

Finally, research 3 results showed that current practice on non-cost score ranges concentrates most of these values (70%) between 0.7 and 0.93. Widening the non-cost range might lead highway agencies to better distinguish the proposers. Future research should identify the reasons that make evaluation teams score in a narrow range. Some of these reasons might be (1) what evaluation criteria are considered. If the criteria themselves do not serve to differentiate, the scores will be similar. (2) How the evaluation criteria are formulated. If evaluation criteria are not specific, the scores will not be consistent. (3) The evaluation team experience. If the team does not have so much experience might be afraid of scoring at the ends of the scale. Future research might survey DOTs to identify these and other reasons. Having a first diagnose, case studies might be conducted to analyze and investigate more specific causes and potential solutions.

Overall, future research should focus its efforts on improving specific aspects of the project delivery systems and procurement methods aiming to improve the design and construction of highway projects.

5.4 Researcher Reflections

This research has been a journey to re-visit old places with new eyes. In the construction industry, we need to keep rethinking our problems and propose research-based practical solutions to remove the inertia that exists in the processes because “this is how things have always been done.”

This research has helped me to take distance from the trees to see the forest. Trees of my fast-pace-industry experience; forest of the general context in which I have been developing my professional career. I have switched my thought from fast to slow. From a fast-immediate-punctual- problem-solving to a reflexive-deep-long-run analysis of the problems.

I realized that both individual and systematic actions need to be done to reach better performance in the design and construction of road and highway projects.

In countries like Chile or Spain, where the standard contracting strategy in highway projects is the traditional approach, there is a need to make decision-makers and society knowledgeable about how project delivery systems and procurement methods influence the project's performance. An individual action as a civil engineer is to strive for making research in this field visible and relevant to ordinary people and decision-makers.

Research on this field might push the transformation of highway agency policies from traditional to more collaborative and effective.

In this regard, there is also a need to foster a collaborative culture. According to Christensen (2016), "change [in an organization's culture] can occur in two ways: as a result of a clear and present crisis or through a managed evolution under a skilled and sophisticated manager." An individual action as a civil engineer is to be the skilled manager (in the university, in the company, in the department of transportation) who is able to introduce and/or propose smoothly small changes that can transform the culture in the long run.

Small changes might push the transformation of highway agency and companies' organizations from silos and individual interest to collaboration and shared benefits.

In research 1, I proposed a cost overrun mitigation measure related to the collaboration of designers and constructors in the design phase. One first approach to consider this measure might be based on the concept of “Bridging.” In design-build projects, “bridging” means that the public owner hires a design professional as a “bridging consultant” who remains involved in the design and construction as an advisor to the owner. In design-bid-build projects, the “bridging consultant” could be a Construction Manager who adds the construction knowledge to the design-phases. Further, road administrators can initiate collaboration with constructors and take advantage of their expertise by requiring alternative technical concepts (ATCs) in the procurement. These ATCs can be focused on optimizing specific design parts produced in the design phase.

These are only two examples that can serve to start talking about steps toward more integration; to move from individual actions to systematic changes.

Transitioning to a collaborative culture is not new, and several countries can be taken as examples in the use of alternative approaches in highway projects. This is the case of the U.S., U.K, New Zealand, Australia, or Finland. It is relevant to conduct research across different countries in order to find synergies. Countries starting to change can learn from the ones that have already accomplished this transition. This is something that I have found very interesting by conducting dual doctoral research between Chile and the US.

The use of collaborative approaches, however, does not mean not having new problems. In this thesis’s research 2 and 3, I addressed one of the challenges that are currently happening in design-build highway projects in the U.S. By conducting theses research, I reinforced my idea of “having a holistic mindset is fundamental.” In other words, an individual action as a civil engineer is to be aware of the “big picture,” so that we can measure the impact in the outcome of each of the process’s steps. This enables us to take the appropriate measures in advance to reach the outcome that we plan.

This mindset needs to be translated from individuals to organizations. Generally, the bigger is the organization, the more segregated are the processes. Although organizations can change suddenly, as I said before, small changes can start the change. Research 1 introduces the measure “high-level point of accountability” as an example of the concept of a “holistic mindset.”. This one point of accountability would be a person that controls the “big picture” of the project and makes sure that schedule and fundings are coordinated to ensure the smooth transition from design to construction in DBB projects. This is a small change that can be introduced now, and that can foster more changes in the future.

In research 2, I analyzed the evaluation criteria included in the Request for Proposals (RFPs). The evaluation criteria represent the metric that highway agencies use to evaluate proposers and decide who is better to develop the work. Previous studies and my research 1’s results showed the relevance of having capable teams to develop the work. However, evaluation criteria, in most cases, are not written with the necessary specificity to make a comprehensive evaluation. Why? Perhaps there is no time to elaborate on specific evaluation criteria for each project. Or, there is not enough knowledge about what is the relevant information that should be required to the proposers. Or, there is not a holistic view under which project’s goals are linked with evaluation criteria, weights and scores, the team selection, the project performance, and ultimately with the project outcomes that should reach the planned project’s goals. The structured approach proposed in research 2 is conceived under this vision. Individual actions defining evaluation criteria under this approach could lead to a systematic change in how Departments of Transportation define evaluation criteria in their procurements.

Finally, the scoring analysis conducted in research 3 makes me link this final part with the middle and the beginning of this reflection section.

With the middle, because it is possible to find synergies between the analysis and results of this research and the Chilean practice on selecting the design team in design-bid-build road projects. The Chilean *Reglamento Para la Contratacion de Trabajos de Consultoria* establishes a best-value procedure to select design and construction supervision teams. The award algorithm is the weighted sum, and the regulation establishes ranges for the weights of cost 0.15 and 0.40. But what is it really happening in practice? What is the trend for the balance between cost and non-cost factors in the selection? Since one of the main reasons for having cost overruns in Chilean road projects is the lack of designers' capabilities, it could be a good idea to follow the research 3 approach in this context.

The link of research 3 with the beginning lies in the idea of "analyzing the weighted sum method does not seem very innovative." It is a commonly used method, and there is not so much to tell about. However, precisely because we are so used to it and it is so easy to use, we are not able to see any problem in it. To keep progressing, it is essential to be able to see old things with new eyes; in both our individual's action perimeter and the agencies' policies framework.

In summary, and as a last thought, I think a good motto for Construction Management research is, what Proust already said in his novel *La Prisonnière,* "The real voyage of discovery consists not in seeking new landscapes but in having new eyes."

References

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

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

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

Adam, A., Josephson, P.-E. B., and Göran, L. (2017). "Aggregation of factors causing cost overruns and time delays in large public construction projects. Trends and implications." *Engineering, Construction and Architectural Management*, 24(3), 393–406.

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

AIA. (2007). *Integrated Project Delivery : A Guide*. *The American Intitute of Architects*.

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

Akcay, C., and Manisali, E. (2018). "Fuzzy decision support model for the selection of contractor in construction works." *Revista de la Construccion*, 17(2), 258–266.

Al-Hazim, N., Salem, Z. A., and Ahmad, H. (2017). "Delay and Cost Overrun in Infrastructure Projects in Jordan." *Procedia Engineering*, 18–24.

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

Alarcón, L. F., Wegmann, A., and Calahorra, M. (2018). "Oportunidades para el mejoramiento de la gestión de proyectos de infraestructura pública en Chile." *Propuestas para Chile, 2017*, 199–239.

Amoatey, C. T., and Ankrah, A. N. O. (2017). "Exploring critical road project delay factors in Ghana." *Journal of Facilities Management*, 15(2), 110–127.

Anderson, S. D., and Russell, J. S. (2001). *Guidelines for warranty, multi-parameter, and best value contracting. NCHRP Report 451*, Transportation Research Board of the National Academies, Washington D.C.

Arditi, D., Akan, G. T., and Gurdamar, S. (1985). "Cost overruns in public projects." *International Journal of Project Management*, 3(4), 218–224.

Asmar, M. El, Lotfallah, W., Whited, G., and Hanna, A. S. (2010). “Quantitative Methods for Design-Build Team Selection.” *Journal 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., and Gonzalez-Cruz, M. C. (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.

Baloi, D., and Price, A. D. F. (2003). “Modelling global risk factors affecting construction cost performance.” *International Journal of Project Management*, 21(4), 261–269.

Barron, F. H., and Barrett, B. E. (1996). “Decision Quality Using Ranked Attribute Weights.” *Management Science*.

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

Bhargava, A., Anastasopoulos, P. C., Labi, S., Sinha, K. C., and Mannering, F. L. (2010). “Three-Stage Least-Squares Analysis of Time and Cost Overruns in Construction Contracts.” *Journal of Construction Engineering and Management*, 136, 1207–1218.

Bordat, C., McCullouch, B. G., Labi, S., and Sinha, K. (2004). *An analysis of cost overruns and time delays of INDOT projects. FHWA/IN/JTRP-2004/7, SPR-2811*.

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

Brierley, G. S., Corkum, D. H., and Hatem, D. J. (2010). *Design-Build Subsurface Projects*. SME.

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.

Cheung, S. O., Yiu, T. W., and Lam, M. C. (2013). “Interweaving Trust and Communication with Project Performance.” *American Society of Civil Engineers*, 139(August), 941–951.

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.

Christensen, C. M. (2016). “What is an Organization’s culture?” *Harvard Business School*, August(399–104), 1–8.

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.

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

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

Commonwealth of Australia. (2008). *National Public Private Partnership Guidelines. Overview*.

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

Creedy, G. D., Skitmore, M., and Wong, J. K. W. (2010a). “Evaluation of Risk Factors Leading to Cost Overrun in Delivery of Highway Construction Projects.” *Journal of Construction Engineering & Management*, 136(5), 528–537.

Creedy, G. D., Skitmore, M., and Wong, J. K. W. (2010b). “Evaluation of Risk Factors Leading to Cost Overrun in Delivery of Highway Construction Projects.” *Journal of Construction Engineering and Management*, 136(5), 528–537.

Creswell, J. W. (2009). *Research design. Qualitative,Quantitative and Mixted Methods Approaches*. SAGE Publications Inc, Thousand Oaks, California.

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. (2014). “Transportation Sector.”

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.

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

DeWitt, S., Yakowenko, G., Bohuslav, T., Ferguson, T., Hoelker, E., Molenaar, K. R., Schiess, G., Smythe, J., Triplett, J., and Wagman, R. (2005). “Management Practices in Canada and Europe.” 1–72.

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.

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

FHWA. (2006a). “Design-Build Effectiveness.” 1307(January).

FHWA. (2006b). *Design-Build Effectiveness*.

FHWA. (2017a). “Briefing On FHWA Innovative Contracting Practices, SEP-14 - Contract Administration - Construction - Federal Highway Administration.” <https://www.fhwa.dot.gov/programadmin/contracts/sep_a.cfm#s1> (May 4, 2019).

FHWA. (2017b). *Alternative Contracting Method Performance in US Highway Construction*. McLean, VA.

Flyvbjerg, B., Holm, M. S., and Buhl, S. (2002). “Underestimating costs in public works projects: Error or lie?” *Journal of the American Planning Association*, 68(3), 279–295.

FMI. (2018a). “Design-Build Utilization.” (June).

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

Forbes, L. H., and Ahmed, S. M. (2011). *Modern Construction:Lean Project Delivery and Integrated Practices.* Taylor & Francis, Boca Raton.

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.

Gibbs, G. R. (2007). “Analyzing qualitative data.” *The Sage qualitative research kit*, SAGE Publications Inc, London.

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.

Gransberg, D. D., and Ellicott, M. A. (1996). "Best value contracting: Breaking the low-bid paradigm." *Transactions of AACE International*, VEC51–VEC56.

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

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

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

Hasanzadeh, S., Esmaeili, B., Nasrollahi, S., Gad, G. M., and Gransberg, D. D. (2018a). "Impact of Owners' Early Decisions on Project Performance and Dispute Occurrence in Public Highway Projects." *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 10(2), 04518004.

Hasanzadeh, S., Esmaeili, B., Nasrollahi, S., Gad, G. M., and Gransberg, D. D. (2018b). "Impact of Owners' Early Decisions on Project Performance and Dispute Occurrence in Public Highway Projects." *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 10(2), 04518004.

Hashem M. Mehany, M. S., Bashettiyavar, G., Esmaeili, B., and Gad, G. (2018). "Claims and Project Performance between Traditional and Alternative Project Delivery Methods." *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 10(3), 04518017.

Hill International. (2016). "Review of WSDOT´s implementation of Design-Build Project Delivery. Final Report." <https://www.wsdot.wa.gov/sites/default/files/2006/01/30/DesignBuild-JTC-FinalReport.pdf>.

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.

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.

Khabisi, J., Aigbavboa, C., and Thwala, W. (2016). "Causes of Cost Overruns in Public Sector Construction Projects in South Africa." *ICCREM*, 1311–1317.

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."

Lam, E. W. M., Chan, Albert P, C., and Chan, D. W. M. (2006). "Barriers to applying the Design-Build procurement method in Hong Kong." *Architectural Science Review*, 49(2), 189–195.

Lorentziadis, P. L. (2010). "Post-objective determination of weights of the evaluation factors in public procurement tenders." *European Journal of Operational Research*, Elsevier B.V., 200(1), 261–267.

Love, P. E. D., Sing, C.-P., Wang, X., Irani, Z., and Thwala, D. W. (2012). "Overruns in transportation infrastructure projects." *Structure and Infrastructure Engineering*, 10(2), 141–159.

Mehany, M., Gad, G. M., and Esmaeili, B. (2017). "Dispute prevention and resolution methods used on public highway projects employing different project delivery methods." *Transportation Research Board 96th Annual Meeting.*, Transportation Research Board of the National Academies, Washington D.C.

Mevada, J., and Devkar, G. (2017). "Analysis of reasons for cost and time overrun in Indian megaprojects." *MATEC Web of Conferences*, 1–10.

Miller, J. B. (2000). *Principles of public and private infrastructure delivery*. Springer Science+Business, New York.

Minchin Jr., R. E., Li, X., Issa, R. R., and Vargas, G. G. (2013). "Comparison of Cost and Time Performance of Design-Build and Design-Bid-Build Delivery Systems in Florida." *Journal* *of Construction Engineering and Management*, 139(10), 04013007.

Minchin Jr, R. E., Li, X., Issa, R. R., and Vargas, G. G. (2013). "Comparison of Cost and Time Performance of Design-Build and Design-Bid-Build Delivery Systems in Florida." *Journal of Construction Engineering & Management*, 04013007(5), 1–5.

Molenaar, K., Harper, C., and Yugar-Arias, I. (2014). *Guidebook for Selecting Alternative Contracting Methods for Roadway Projects : Project Delivery Methods , Procurement Procedures , and Payment Provisions*. Boulder.

Molenaar, K., and Johnson, D. (2003a). "Engineering the procurement phase to achieve best value." *Leadership and Management in Engineering*, July(3), 137–142.

Molenaar, K. R., and Johnson, D. E. (2003b). "Engineering the Procurement Phase to Achieve Best Value." *Leadership and Management in Engineering*, 3(3), 137–141.

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.

MOP Chile. (2016). *Informe de gestión proyectos terminados*.

Morledge, R., and Smith, A. (2013). *Building Procurement*. (J. Wiley, ed.), ProQuest Ebook Central.

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.

OECD/ITF. (2013). "Spending on transport infrastructure 1995-2011. Trends, policies, data." <https://www.itf-oecd.org/sites/default/files/docs/13spendingtrends.pdf> (May 15, 2019).

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.

Pakkala, P. (2002). *Innovative Project Delivery Methods for Infrastructure - An International Perspective*. Finnish Road Enterprise, Helsinki.

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).

Park, J., and Kwak, Y. H. (2016). “Design-Bid-Build (DBB) vs. Design-Build (DB) in the U.S. public transportation projects: The choice and consequences.” *International Journal of Project Management*, Elsevier Ltd, APM and IPMA, 35(3), 280–295.

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

Plusquellec, T., Lehoux, N., and Cimon, Y. (2017). “Design-Build and Design-Bid-Build in Construction- a Comparative Review.” *IGLC*, 35–43.

PMI. (2014). *Study on project schedule and cost overruns*. *Project Management Institute*, India.

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

Rosenfeld, Y. (2014). “Root-Cause Analysis of Construction-Cost Overruns.” *Journal of Construction Engineering and Management*, 140(401), 30–39.

Ruparathna, R., Asce, S. M., and Hewage, K. (2015). “Review of Contemporary Construction Procurement Practices.” 31(3), 1–11.

Saaty, T. (1980). *The Analytic Hierarchy Process*. Mac Graw Hill, New York.

San Cristóbal, J. R. (2011). “Contractor Selection Using Multicriteria Decision-Making Methods.” *Journal of Construction Engineering and Management*, 138(6), 751–758.

Sanvido, V., and Konchar, M. (1999). *Selecting project delivery systems: comparing design build, design-bid-build and construction management at risk (CII)*.

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.

Schöttle, A., and Arroyo, P. (2017). “Comparison of Weighting-Rating-Calculating, Best Value, and Choosing by Advantages for Bidder Selection.” *Journal of Construction Engineering and Management*, 143(8), 05017015.

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.

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.

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.

Shane, J. S., Molenaar, K. R., Anderson, S., and Schexnayder, C. (2009). "Construction Project Cost Escalation Factors." *Journal of Management in Engineering*, 25(4), 221–229.

Shrestha, P. P., Migliaccio, G. C., O'Connor, J. T., and Gibson, G. E. (2007). "Benchmarking of Large Design-Build Highway Projects: One-to-One Comparison and Comparison with Design-Bid-Build Projects." *Transportation Research Record: Journal of the Transportation Research Board*, 1994(1), 17–25.

Shrestha, P. P., O'Connor, J. T., and Gibson, G. E. (2012). "Performance Comparison of Large Design-Build and Design-Bid-Build Highway Projects." *Journal of Construction Engineering and Management*, 138(1), 1–13.

Sohu, S., Abdullah, A. H., Nagapan, S., Fattah, A., Ullah, K., and Kumar, K. (2017). "Contractors perspective for critical factors of cost overrun in highway projects of Sindh, Pakistan." *AIP Conference Proceedings*, 1892(080002), 1–6.

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

Sullivan, J., El Asmar, M., Chalhoub, J., and Obeid, H. (2017). "Two Decades of Performance Comparisons for Design-Build, Construction Manager at Risk, and Design-Bid-Build: Quantitative Analysis of the State of Knowledge on Project Cost, Schedule, and Quality Alternative Project Delivery in Construction." *Journal of Construction Engineering and*

Management, 143(6).

Tala González, N. I. (2015). "Tesis Magister. Identificación de causas que generar modificación de plazos y costos en contratos de ejecución de obras públicas entre los años 2005 y 2015 (Chile)." Pontificia Universidad Católica de Chile.

The Economist. (2017). "Efficiency eludes the construction industry." *Business*, <https://www.economist.com/business/2017/08/17/efficiency-eludes-the-construction-industry>.

Touran, A., Gransberg, D. D., Molenaar, K. R., Ghavamifar, K., Mason, D. J., and Fithian, L. A. (2008). *TCRP. Document 41: Evaluation of Project Delivery Methods*. Transportation Research Board of the National Academies, Washington, D.C.

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

Tran, D. Q., Diraviam, G., and Minchin Jr, R. E. (2018b). "Performance of Highway Design-Bid-Build and Design-Build Projects by Work Types." 144(2), 1–9.

Tran, D. Q., Molenaar, K. R., and Kolli, B. (2017a). "Implementation of best-value procurement for highway design and construction in the USA." *Engineering, Construction and Architectral Management*, 24(5), 774–787.

Tran, D. Q., Nguyen, L. D., and Faught, A. (2017b). "Examination of communication processes in design-build project delivery in building construction." *Engineering, Construction and Architectral Management*, (1), 113–131.

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

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

Verweij, S., van Meerkerk, I., and Korthagen, I. A. (2015). "Reasons for contract changes in implementing dutch transportation infrastructure projects: An empirical exploration." *Transport Policy*, Elsevier, 37, 195–202.

Verweij, S., and Meerkerk, V. (2015). "Reasons for Contract Changes in Implementing Dutch Transportation Infrastructure Projects : An Empirical Exploration Reasons for Contract Changes in Implementing Dutch Transportation Infrastructure Projects : An Empirical Exploration Reasons for Contract." *Transport Policy*, 37, 195–202.

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

El Wardani, M. A., Messner, J. I., and Horman, M. J. (2006). “Comparing Procurement Methods for Design-Build Projects.” *Journal of Construction Engineering and Management*, 132(3), 230–238.

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

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.

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.

World Bank. (2014). *Source of Time and Cost Over-runs in Roadworks Projects – Pilot Study*. World Bank Group, Sri Lanka.

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.

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

Yu, W., and Wang, K.-W. (2012). “Best Value or Lowest Bid? A Quantitative Perspective.” *Journal of Construction Engineering and Management*, 138(1), 128–134.

Appendices

A. Interviewees characteristics (Research 1)

The research aimed to identify the primary reasons for cost overruns in the design, the procurement, and the construction phases in Chilean road Projects. To this end, the interviewees were selected based on three criteria that aimed to consider the main factor that can influence the interviewees' perception of the main reasons for cost overruns (Table 20).

Table 20 Criteria for selecting interviewees

Criteria	Rationale
Years of experience	More than five (5) years of experience in the management of design and/or construction in road projects[a] in Chile. More experience provides more sound responses.
Work experience location	The author looked for representation from the North, Middle, and South Chilean regions since the context varies substantially between them, and this might influence the types of reasons for cost overruns that mainly occur.
Sector	The author looked for representation from both the Public and Private Sectors since the approach to cost overruns might differ substantially.

[a] Rehabilitation/Conservation Projects not considered.

Table 21 shows the interviewee's position and sector.

Table 21, Interviewees. Position and Sector

	Entrevistas. Motivos Modificaciones de obra		
	Cargo o Departamento	**PUBLICO/PRIVADO**	**AMBITO**
1	Jefe de proyecto region Sur	PUBLICO	FUNCIONARIO MOP
2	Jefe Asistencia Tecnica Region Central	PRIVADO	INSPECCIÓN TECNICA
3	Administrador Obra V region	PRIVADO	CONSTRUCTOR
4	Jefe oficina Tecnica	PRIVADO	CONSTRUCTOR
5	Ex director regional mop	PUBLICO	FUNCIONARIO MOP
6	Ex funcionario MOP	PUBLICO	FUNCIONARIO MOP
7	Supervisor civil de obras	PRIVADO	INSPECCIÓN TECNICA
8	Seremi mop 5 región	PUBLICO	SEREMI
9	Jefe Contratos	PUBLICO	FUNCIONARIO MOP
10	Profesional experiencia en ITO	PRIVADO	INSPECCIÓN TECNICA
11	Responsable en Vialidad	PUBLICO	FUNCIONARIO MOP
12	Subdirección Obras	PUBLICO	FUNCIONARIO MOP
13	Dpto de Construcción	PUBLICO	FUNCIONARIO MOP
14	Dpto de Gestión Vial	PUBLICO	FUNCIONARIO MOP
15	IF obra en Región Metropolitana	PUBLICO	FUNCIONARIO MOP
16	IF obra en Región Metropolitana	PUBLICO	FUNCIONARIO MOP
17	Administrador obra Región Metropolitana	PRIVADO	CONSTRUCTOR
18	IF obra en Región Norte	PUBLICO	FUNCIONARIO MOP
19	Jefe Asistencia Tecnica Región Norte	PRIVADO	INSPECCIÓN TECNICA
20	Administrador obra Región Norte	PRIVADO	CONSTRUCTOR
21	IF obra en Región Norte	PUBLICO	FUNCIONARIO MOP
22	IF obra en Región norte	PUBLICO	FUNCIONARIO MOP
23	Jefe Asistencia Tecnica Región Norte	PRIVADO	INSPECCIÓN TECNICA
24	Administrador obra Región Norte	PRIVADO	CONSTRUCTOR
25	IF obra en Región Norte	PUBLICO	FUNCIONARIO MOP
26	IF obra en Región Sur	PUBLICO	FUNCIONARIO MOP
27	Jefe Asistencia Tecnica Región Sur	PRIVADO	INSPECCIÓN TECNICA
28	Administrador obra Región Sur	PRIVADO	CONSTRUCTOR
29	IF obra en Región Sur	PUBLICO	FUNCIONARIO MOP
30	Jefe Asistencia Tecnica Región Sur	PRIVADO	INSPECCIÓN TECNICA
31	Administrador obra Región Sur	PRIVADO	CONSTRUCTOR
32	IF obra en Región Sur	PUBLICO	FUNCIONARIO MOP
33	Jefe Asistencia Tecnica Región Sur	PRIVADO	INSPECCIÓN TECNICA
34	Administrador obra Región Sur	PRIVADO	CONSTRUCTOR
35	IF obra en Región Central	PUBLICO	FUNCIONARIO MOP
36	Administrador obra Región Central	PRIVADO	CONSTRUCTOR
37	IF obra en Región Central	PUBLICO	FUNCIONARIO MOP
38	Jefe Asistencia Tecnica Región Central	PRIVADO	INSPECCIÓN TECNICA
39	Administrador obra Región Central	PRIVADO	CONSTRUCTOR
40	IF obra en Región Central	PUBLICO	FUNCIONARIO MOP
41	Administrador obra Región Central	PRIVADO	CONSTRUCTOR

B. CONTENT ANALYSIS DETAIL (RESEARCH 1)

FLOW CHART

The research followed the process of inductive content analysis, as shown in Figure 19.

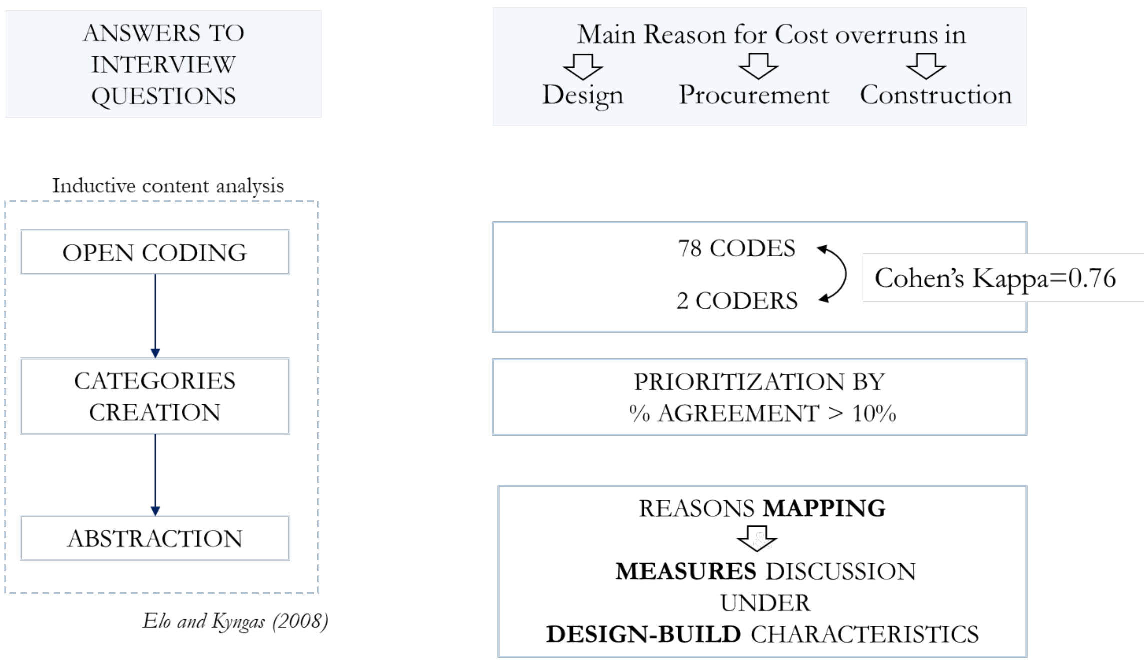


Figure 19 Inductive content analysis flow chart for Research 1

CODE-BOOK

The author created a codebook that served to codify the interview by two different coders. The inter-reliability between them was calculated using Cohen's Kappa metric.

Table 22 constitutes the codebook used to codify the interviews' responses.

Table 22 Codebook. Research 2

	Code	Definition
1	Errors & Omissions in design	This code refers to the incomplete design, vagueness in the design, deficient design, inconsistent design, mistakes, and/or incoherence in the bill of quantities and/or prices.
2	Lack of constructability	Due to the lack of interaction between designer & constructor, the design is not optimized for construction, and the definition of the construction stages and their durations are not well defined.
3	Inadequate designer capabilities	It includes different aspects associated with the lack of capabilities that the designer has to develop the design contract.
4	Poor design skills and design management	The design team lacks good coordination; it is too theoretical, it has not proper capabilities
5	Lack of a designer's experience	The designer has not enough experience to develop the project properly
6	Lack of designers local knowledge	The designer does not know specific information on the area where the project is being designed.
7	Lack of coordination between the areas	There is no coordination between design disciplines
8	Lack of professionals and resources to develop the design	There are not enough professionals to develop the work, or they have more work than they can assume
9	Inadequate quality control in design	It refers to a deficient design quality control system.
10	Inadequate owner supervision in design	It refers to the excess of reviewing during the design phase.
11	Lack of proper basic engineering	It refers to the baseline information that should be gathered to develop the project. Specifically, topography, geotechnics, hydraulics, traffic, and register of current infrastructures in the area of the project.
12	Poor Topography	The topographic information needed to develop the work is not sufficient or has errors.
13	Poor soils identification and/or geotechnical studies	The soils identification and/or geotechnical studies needed to develop the work are not sufficient or has errors.
14	A poor general baseline of the project's area conditions	The general information of the project's area is vague, inaccurate, or contains mistakes.
15	Lack of owner flexibility in the conceptual design	The owner does not allow flexibility in the conceptual scope nor alternative solutions
16	Lack of owner independent review in design	The public administrator does not have an independent reviewer for the design

17	Long time for review & approve the design	It refers to the long time that is required to approve the projects, both by the Ministry of Public Works and by other related Public Administrations.
18	Short time to develop engineering	It refers to the short time defined to develop the design documents
19	Poor design scope definition	It refers to the lack of clarity in the definition of the design scope. This means that the requirement for the designers to develop their work might not be clear.
20	Low budget for developing design contracts	It refers to the fact low economic resources are allocated to develop design contracts.
21	Owner's departments of design & construction do not talk between them in design phase	In the public administration, the Design team do not have communication with the Construction team
22	Do not check the project in situ when the design is finished	The designer does not represent "in situ" the geometrical design when it is finished; therefore, there is not a real check between the design and the real conditions
23	Lack of design revision & update	It refers to review the projects previously to include them in the bidding documents. This revision might include checking the design in situ, review and update the design documents, among other reasons.
24	Lack of modernization of the bidding system	The bidding system is not modern.
25	Long time to award the contract	It takes a long time to award the construction contract.
26	Lack of proper information exchange in the procurement	The public administration do not properly clarify the proposer's questions
27	Budget cut	Constrains the budget to adapt it to the available resources
28	High weight of price in contrast with technical factors	The selection process prioritizes more the price than other technical factors
29	Inadequate constructor selection	The selection of the constructor company is not adequate.
30	Inaccurate proposals	The proposal submitted by the construction company is not detailed.
31	Proposer's economic proposal over the official budget	The economic proposal is over the administration's estimated budget.
32	Remove part of the design	Some parts of the design are removed from the documents before the construction is awarded.
33	Restricted budget to study the RFP	The budget allocated to study the RFP and prepare the proposal is limited.
34	Short time to study & clarify RFP	The time to study the RFP and develop the proposal is short.

35	Vague/Incoherent/with errors Request For Proposal (RFP)	It refers to the poor quality of the RFP
36	Bidding process with inefficiencies	It has no administrative order, centralizes the decision making, uses a long time in the administrative order, instead of technical. Furthermore, it has no clarity regarding the scope of the work.
37	Long time between design & construction	There is a long time between the design finishes and the construction starts
38	Change in soil foundation conditions	The foundation soils are different than the ones included in the design
39	Inadequate constructor capabilities & interests	It refers to construction company capabilities to develop project construction.
40	Poor construction Management	It refers to poor management in general.
41	Construction practices. Poor management and mistakes	It refers to improper constructive practices, lack of resources to develop the construction & mistakes during the execution
42	Poor construction Problem solving	It refers to not being able to visualize problems and solving them in advance
43	Inadequate team skills	It refers to inadequate teams to manage the construction
44	Constructor working for his benefit	It refers to the construction company driven by their economic benefit
45	Change in construction methodologies	It refers to modifications in the construction methodology
46	In situ adjustments to real conditions	The design is not coherent with the real conditions when the project is going to be built; thus, adjustments are needed.
47	Deficient owner supervision in construction	Characteristics of the public administrator's supervision team that could include: extending the revision time in their benefit, lack of resources, inefficiency performing the work, lack of experience, lack of decision-making skills.
48	Delays related to developing project changes	It refers to the increase in time associated with the development of project changes during construction.
49	Different owner's technical criteria in design and construction	The public administrator that reviews the construction is different than the one that reviewed the design. This fact generates criterion differences at the moment of construction.
50	Increase in quantities	It refers to the increase in quantities originally included in the design. This increment might occur because "in situ adjustments," "necessity of new items," "errors and omissions in the design," among other reasons.
51	Interferences with infrastructures not detected	It refers to interferences with non-detected (in design) infrastructures
52	Lack of coordination with other contracts & third parties in construction	It refers to a lack of coordination with other contracts that might be in the same project area

53	Need to include new items during construction	It refers to include new items to the project during the construction phase.
54	Political factors	It refers to how the political environment can influence construction decisions.
55	Project outdated	The project is not updated when it is going to be built.
56	Weather/Climate	Weather conditions
57	Unforeseen events	Contingencies. Unexpected events during construction
58	Slow process for authorizing project changes in construction	The Chilean process to authorize project changes in construction takes a long time. While the change is not approved is not possible to progress the construction in part affected by the modification.
59	Causes of force majeure	Natural events, such as earthquakes or special weather conditions, occur.
60	Community/Third party requirements	The community affected by the construction asks for including more items.
61	Construction mistakes	The construction company incurs in mistakes that need to be solved
62	Poor/Inadequate construction scheduling	It is referred to as deficient constructor scheduling and also inadequate owner scheduling
63	Problems with local communities/third parties	Problems generated by third parties that might be affected by the construction
64	Project changes	It refers to project changes that need to be performed during the construction.
65	Problems related to obtaining materials for construction	It refers to the lack of proper study of materials in design and also problems that the constructor might face to obtain these materials during the construction.
66	Lack of risk analysis	It refers to the lack of risk analysis: identification, allocation, follow-up during the life of the project.
67	Change in the regulation	The technical regulation that was used to develop the design change before starting the construction; thus, the design needs to be updated.
68	Change of traffic conditions	The traffic conditions that were used to develop the design change before starting the construction: thus, the design needs to be updated.
69	Changes in the project area	The ground and area conditions may change since the project is finished and the construction is started
70	Bureaucracy	It means that administrative procedures and activities are required.
71	Corruption	It is dishonest or fraudulent conduct by those in power, typically involving bribery.
72	Environmental issues	It refers to different issues associated with the environment that can occur in different phases of the project delivery. This might include but not limited: archeologic findings in construction, mitigation measures not considered in the design, improper development of community outreach in design or/and in construction, Environmental impacts sub estimated
73	Lack of owner involvement & experience	The public administrators in charge of design and construction review are not engaged with the projects, or they do not have enough experience
74	Not considering all the stakeholders	It refers to not considering all the stakeholders that could be affected by the project.

75	Owner's decisions	It refers to the owner's decision in general and/or the lack of proper owner's decision
76	Right of Way issues	It refers to problems in the right of way process and problems in having the land free for the construction use
77	Utility relocation issues	It refers to the conditions in construction to relocate the utilities (coordination with the companies, delays).
78	Social profitability limitations	In Chile, public projects are required to have social profitability, meaning that the project budget (the total investment) is compensated by the social benefit that is obtained when the project is built. Due to this fact, sometimes the scope of design projects is reduced to accomplish this requirement.

DATA AVAILABILITY STATEMENT

The content analysis was developed using the software dedoose (https://www.dedoose.com/). The interviews' excerpts, as well as the codification, can be checked in dedoose's platform upon request to the author.

C. Summary of Request for proposals (Research 2)

Table 23 summarizes the RFPs used in the analysis, including the number of RFP and evaluation criteria type per state. Initially, the analysis considered the 40 states that have a yearly expenditure on highway projects higher than US$ 1 million (>1,000.000 $). From the 40 projects, 11 did not have full authorization for using design-build; therefore, they were not considered.

From the remaining 29 states, eight (8) did not have any information available. Thus, the analysis considered RFPs from 21 states.

Table 23 RFPs analyzed by State and type of evaluation criteria

States	Sample RFP	Evaluation criteria Types A	B	C	Comments	
California	4			11	Same formulation for all the projects	Evaluation Criteria. Appendix E
New York	5	4		34	Quality Evaluation factors. Appendix A	
Texas	5		23	3	Evaluation Factors. Sections 5.3.1-2/5.2	
Florida	8			34	Same formulation for all the projects	Evaluation criteria. Sections C?VII
Illinois		-	-	-	Not specifically authorized	
Ohio	2	2	53		Evaluation criteria. Section 4	
Pennsylvania		-	-	-	Not specifically authorized	
Georgia	3		12		Same formulation for all the projects	Evaluation criteria. Section 5.2 & Exhibit C,E
Iowa					Not specifically authorized	
Kentucky	3		47		Same formulation for all the projects	Evaluation criteria. Technical proposal evaluation
Maryland	10		30		Same formulation for all the projects	Evaluation factors. Section Technical proposal
Massachusetts					No information available	
Michigan	1			3	From Design-Build guidelines. Pag 21	
Minnesota	6	5		12	Evaluation of technical proposal. Section 5 & 4.2	

New Jersey		-	-	-	Not specifically authorized	
North Carolina	5			35	Same formulation for all the projects	Evaluation criteria. Section "Evaluations"
Oklahoma					No information available	
Virginia	4			16	Same formulation for all the projects	Evaluation factors. Section 5
Washington	10	6	27	0	Proposal evaluation process (see state's table)	
Wisconsin		-	-	-	Not specifically authorized	
Alabama		-	-	-	Not specifically authorized	
Alaska	1		1		Design-Build guide. Page 5-3	
Arizona	1			4	Evaluation criteria. Section C-I	
Arkansas					No information available	
Colorado	4	1.5	7	5.5		
Connecticut	2		23		Evaluation criteria. Chapter 8	
Indiana		-	-	-	Authorized. Certain limitations	
Kansas		-	-	-	Authorized. Certain limitations	
Louisiana	6		62		Same formulation for all the projects	Evaluation factors and criteria. Section 6
Mississippi	7			38	Same formulation for all the projects	Section: criteria for scoring
Missouri					No information available	
Nebraska					No information available	
Nevada					No information available	
New Mexico		-	-	-	Authorized. Certain limitations	
North Dakota		-	-	-	Authorized. Certain limitations	
Oregon					No information available	
South Carolina	10			37		
Tennessee					Authorized. Certain limitations	
Utah	1	4			Instruction to proposer’s guidelines. Page 27	
West Virginia					No information available	
	98	**22.5**	**285**	**233**		
	540	4%	53%	43%		

D. Content Analysis Detail (Research 2)

Flow Chart

The research followed the process of inductive content analysis, as shown in Figure 20.

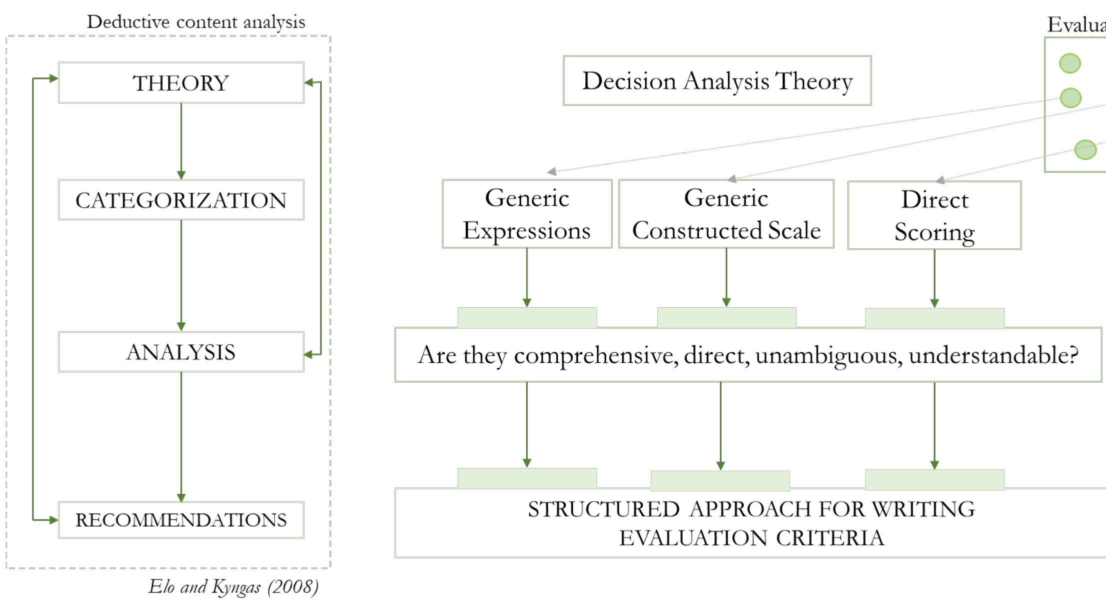


Figure 20 Deductive content analysis flow chart for Research 2

Data Availability Statement

All data that support the findings of this research are available from the author upon reasonable request.

E. Variability & Outliers Analysis (Research 3)

Variability Analysis

In research 3, the author analyzed the variability of the cost and non-cost scores associated with projects' scope and projects' location using the Mann-Whitney U Test.

The Mann-Whitney U test was used because the data were not normally distributed. The analysis was performed using SPSS software.

The project scope variability analysis was conducted using a subset of 127 projects from 13 states, accounting for 36.3% of all the projects considered. These projects provided 368 cost scores and 399 non-cost scores. The sample size in both cases represented the complete sample (1.158 cost scores and 882 non-cost scores) with a 95% confidence level and a 5% margin of error. The results of the Mann Whitney U test showed that there was no statistically significant difference among projects´ scopes for both non-cost scores (Table 24) and cost scores (Table 25).

Table 24 Mann Whitney U Test. Project's Scope. Non-Cost Scores

Group 1	Group 2	P-value
Bridge	Interchange	0,204 (>0.05)
Bridge	Highway	0,954 (>0.05)
Bridge	Several Types	0,208 (>0.05)
Interchange	Highway	0,290 (>0.05)
Interchange	Several types	0.01
Highway	Several types	0,099 (>0.05)

(*) The result of the comparison between "interchange" and "several types" was analyzed under other data, such as the proportion of projects of each type and the box plot graphs of each type (Figure 21, Figure 22). The general approach suggested that there is not enough evidence to determine that interchange scope or several types' scope should be considered independently.

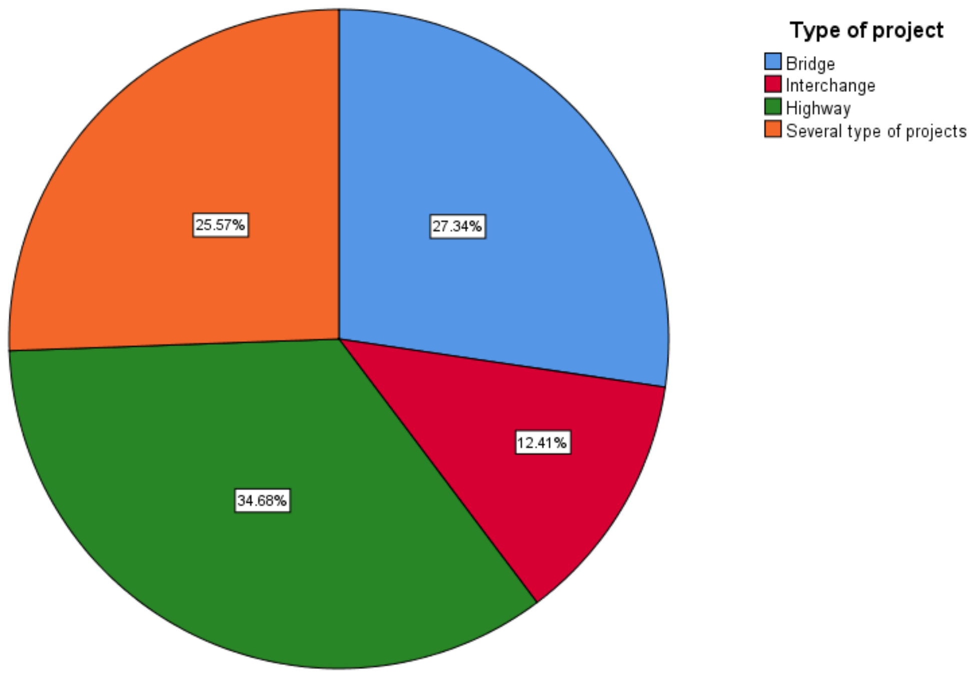


Figure 21 Non-Cost proportion per project's scope type

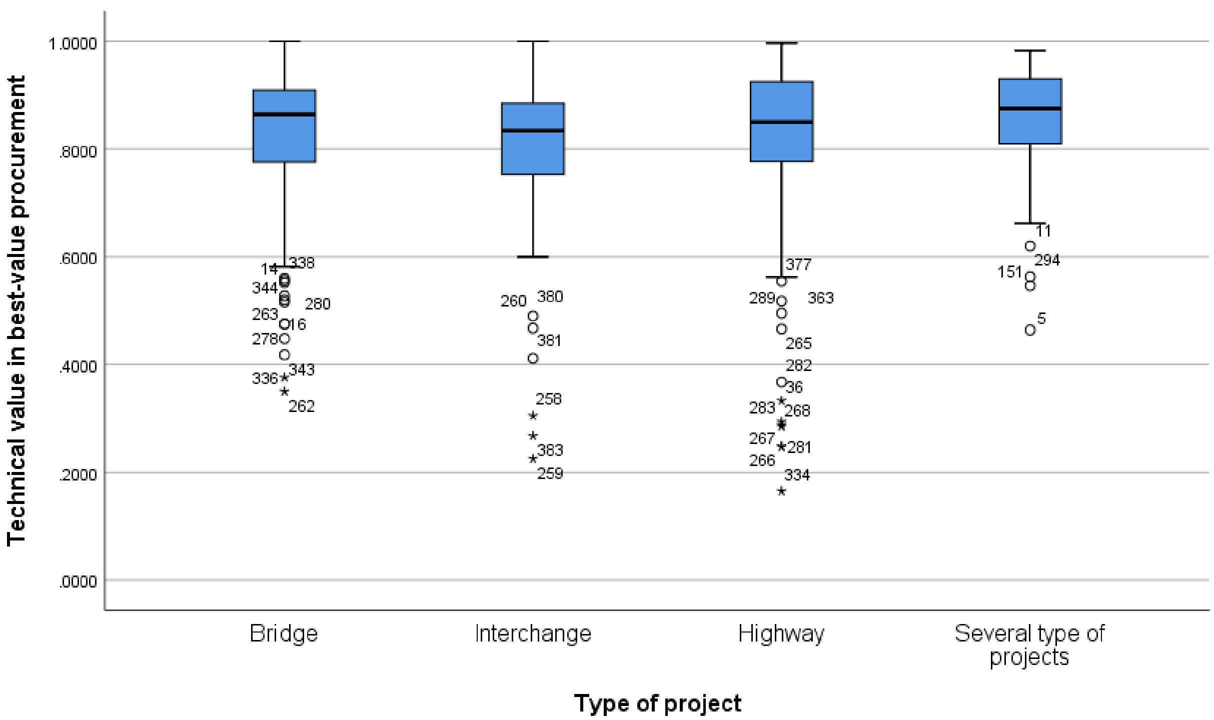


Figure 22 Box-plot of non-cost scores per project's scope type.

Table 25 Mann Whitney U Test. Project's scope. Cost Scores

Group 1	Group 2	P-value
Bridge	Interchange	0,799 (>0.05)
Bridge	Highway	0,563 (>0.05)
Bridge	Several Types	0,911 (>0.05)
Interchange	Highway	0,786 (>0.05)
Interchange	Several types	0,762 (>0.05)
Highway	Several types	0,352 (>0.05)

The Mann Whitney U test was also used to check whether the cost and non-cost scores data from Florida could be considered similar to the dataset comprised of all the other states.

In the non-cost score set, Florida provided 420 data points, whereas all the other states provided 399. In the case of cost scores, Florida provided 415 data, whereas the other states contributed with 715 data points.

The results of the Mann Whitney U test showed that there was no statistically significant difference among Florida and the other states for both non-cost scores (Figure 23) and cost scores (Figure 24).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Technical Score is the same across categories of Non-Florida Stae.	Independent-Samples Mann-Whitney U Test	.577	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,050.

Figure 23 Mann Whitney U Test. Florida and other states. Non-Cost Scores. SPSS output

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Cost Score is the same across categories of FloridaORnoFlorida.	Independent-Samples Mann-Whitney U Test	,154	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Figure 24 Mann Whitney U Test. Florida and other states. Cost Scores. SPSS output

OUTLIERS ANALYSIS

In regards to the outliers, Figure 25 and Figure 26 shows the number of outliers per state.

OUTLIERS. COST SCORES
1% total (1158)

Scores lower than 0.58

Cost bids that were 1.7 times more expensive than the lowest bidder

State	# of outliers	% total	# projects	Comments
Florida	7	0.60%	5	2 project with 2 outliers each
Virginia	4	0.35%	3	1 project with 2 outliers
Georgia	2	0.17%	2	
Idaho	1	0.09%	1	
North Carolina	1	0.09%	1	
Mississippi	1	0.09%	1	
	16			

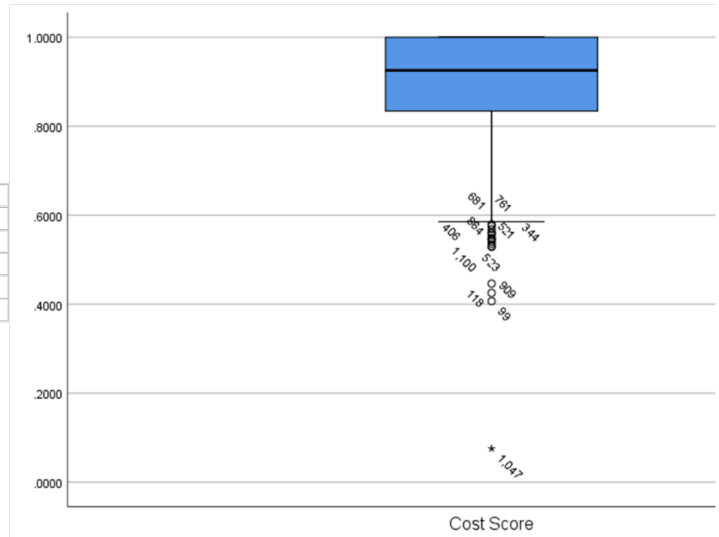


Figure 25 Cost Scores outliers

OUTLIERS. NON-COST SCORES *6% total (822)*

Scores lower than 0.66

State	# of outliers	% total	Time procur.	Comments	
South Carolina	17	2.07%	2014-2016	2 projects with 4 outliers & 3 with 3	
Washington	12	1.46%	2010-2015 (10)	2018-2019 (2)	
Georgia	6	0.73%	2015 (1)	2017-2018 (5)	3 from 1 project
California	4	0.49%	2011-2012		
Virginia	3	0.36%	2015-2016		
Kentucky	2	0.24%	2013		
Minnessota	2	0.24%	2007		
Mississippi	2	0.24%	2011-2013		
Florida	2	0.24%	-		
Louisiana	1	0.12%	2010		
Texas	1	0.12%	2018		
	52				

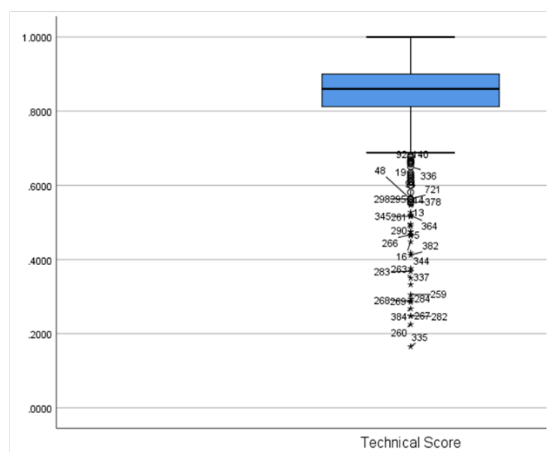


Figure 26 Non-Cost Scores Outliers

Data availability Statement

All data that support the findings of research 3 are available from the author upon reasonable request.

F. THE GOODNESS OF FIT GRAPHS (RESEARCH 3)

The distribution that best-fit cost scores were the Beta distribution, whereas the one that best fit the non-cost scores were the Gumbel distribution. Figure 27, Figure 28, and Figure 29 represent the cost scores and beta distribution in the form of the probability density, cumulative ascending, and p-p graphs.

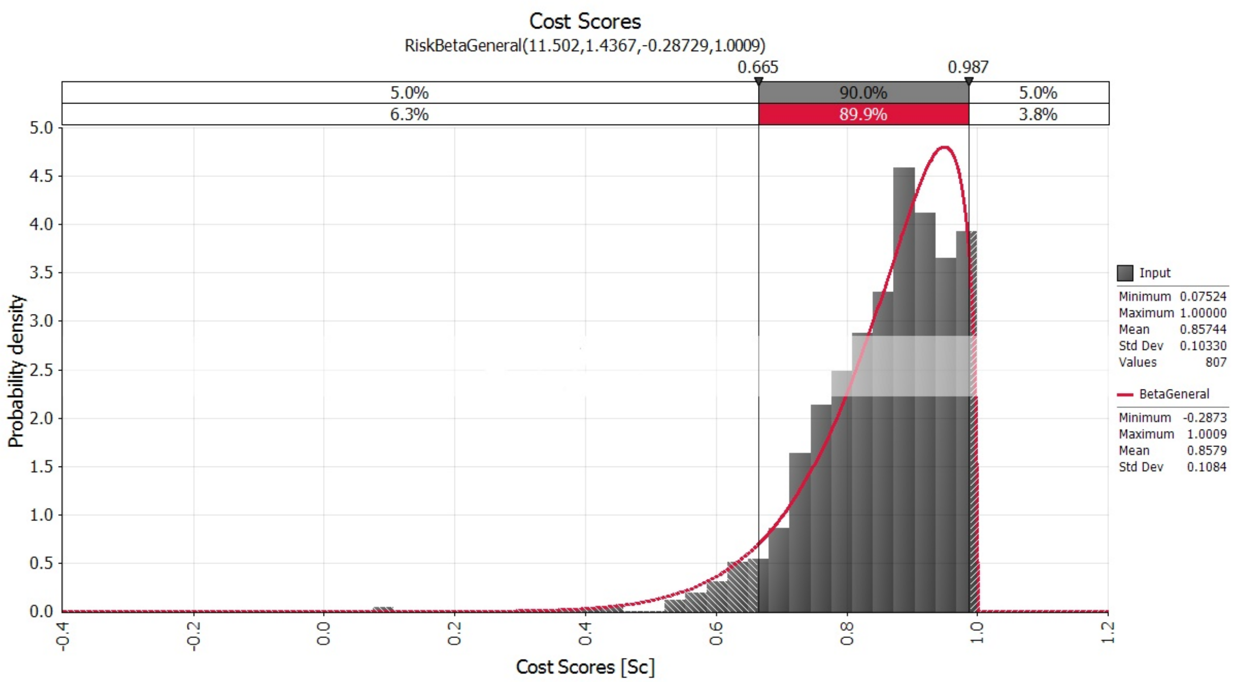


Figure 27 Probability Density. Cost Scores & Beta Distribution

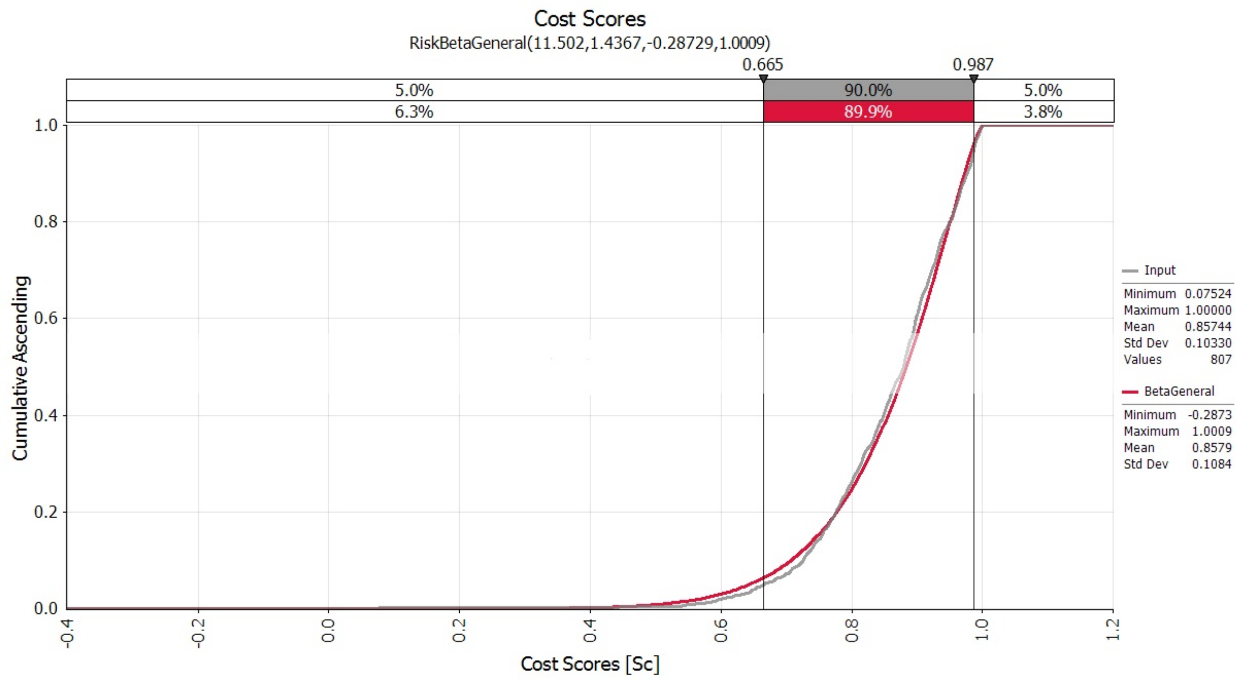


Figure 28Cumulative Ascending Graph. Cost Scores & Beta Distribution

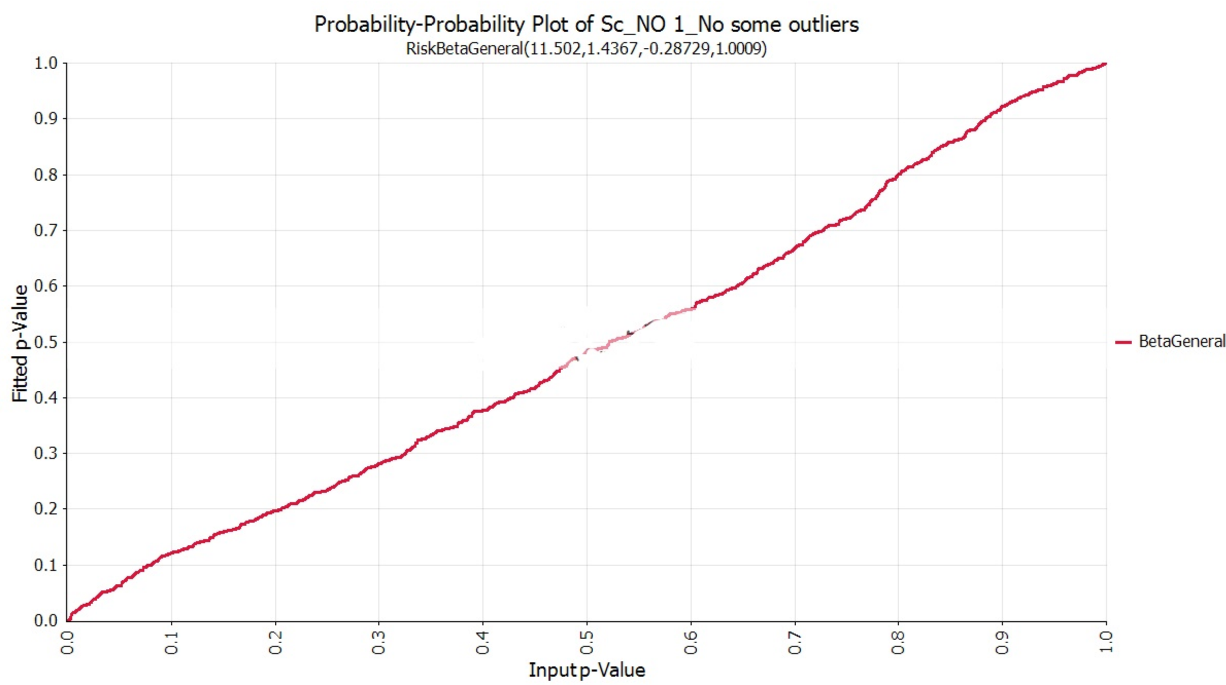


Figure 29 P-P Graph. Cost-Scores & Beta Distribution

Figure 30, Figure 31, Figure 32 represents the cost scores and beta distribution in the form of the probability density, cumulative ascending, and p-p graphs.

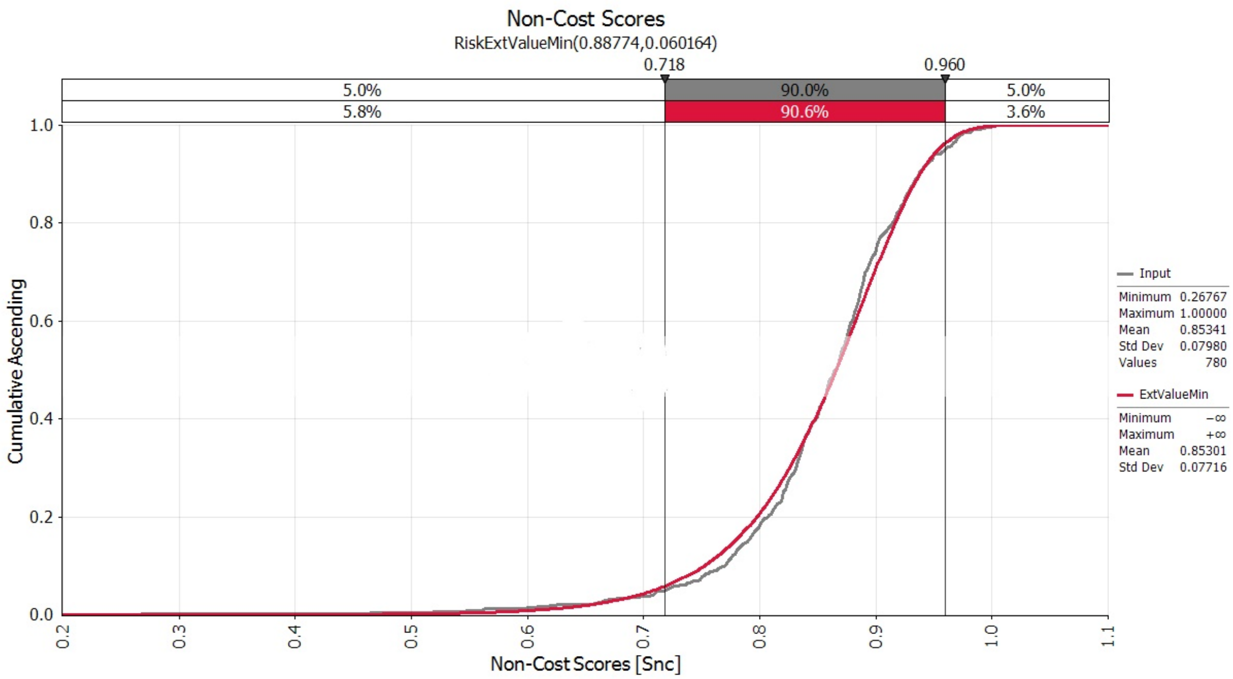


Figure 30 Probability Density. Non-Cost Scores & Gumbel Distribution

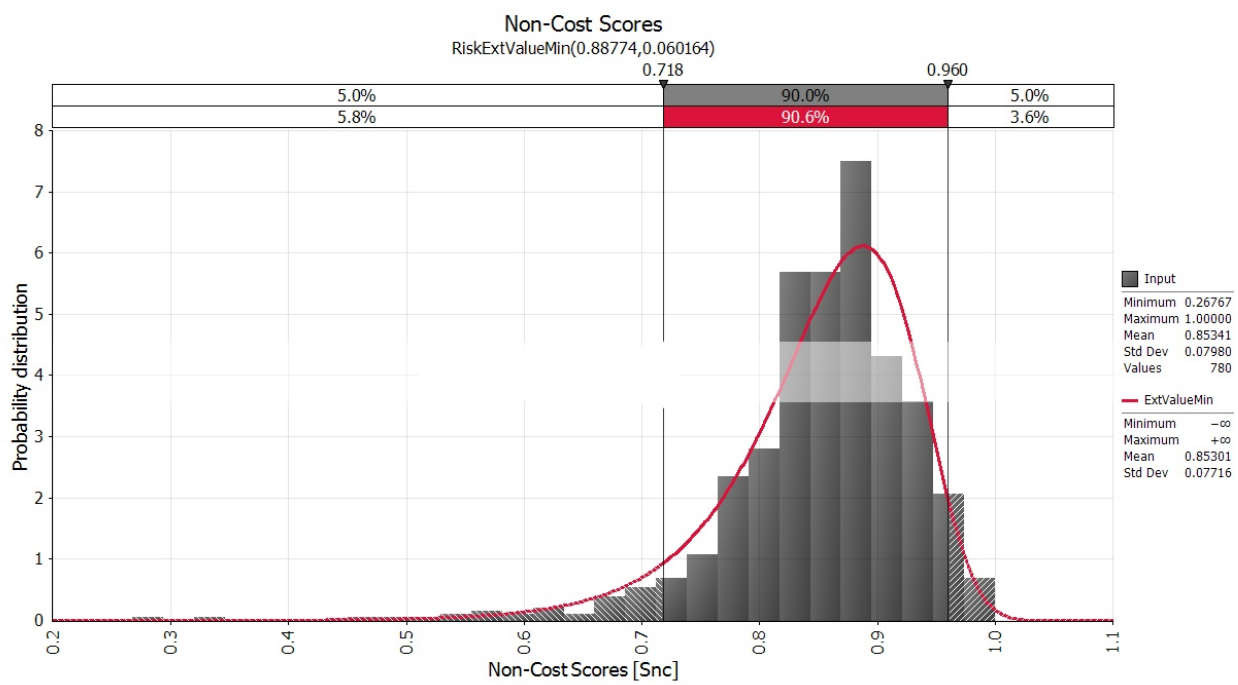


Figure 31 Cumulative Ascending Graph. Non-Cost Scores & Gumbel Distribution

Figure 32 P-P Graph. Non-Cost-Scores & Gumbel Distribution