RESEARCH STUDY

STRATEGIES TO WIN A/E DESIGN PROPOSALS: Which Evaluation Criteria Best Differentiate between Competitors?

by
Simplar

Abstract

Although the procurement of architectural and engineering (A/E) consultants has traditionally followed Qualifications-Based Selection (QBS), owners have begun to utilize cost evaluations with increasing regularity. A/E professional associations have widely viewed this as a threat to their profession and have established guidelines stating that owners should limit the weight of cost evaluations and use a two-envelope system to ensure an unbiased evaluation. This study investigated the application of these guidelines within 122 publicly procured A/E projects across North America. Selection outcomes roughly mimicked QBS, such that the highest qualified consultant was selected in the majority of cases; however, lower pricing did drive a minority of selections. The study also investigated the owner evaluation scores received by each individual consultant proposal within the data set (n = 804). Results contributed an understanding of the evaluation criteria that best differentiate between competing proposals, which informs owners in optimizing their evaluation schemes and guides consultants in their proposal development efforts. Finally, analysis revealed that greater consultant qualifications have no correlation with higher design cost, which contradicts owner-held perceptions that QBS results in higher costs.

Research Details

INTRODUCTION

Qualifications-Based Selection (QBS) has been the predominant procurement method for architectural and engineering (A/E) services for more than 50 years (Chinowsky and Kingsley 2009). The Brooks Act, which was adopted by Congress in 1972, requires all federal projects to procure A/E services on the basis of qualifications and then negotiate a fair and reasonable cost with the selected A/E consultant. According to the American Council of Engineering Companies (ACEC), 46 of 50 states have emulated the federal government by adopting their own QBS laws, oftentimes called “miniBrooks Acts.”

Although almost all states have what is considered a miniBrooks Act at the core of their procurement law, the extent to which this core has exceptions varies widely from state to state. As states continue to integrate alternative delivery strategies (such as design build, construction manager at risk, and integrated project delivery), they are presented with the opportunity to adjust procurement law provisions, which has tended to weaken QBS procurement requirements (Chinowsky and Kingsley 2009). For public entities that are further removed from the federal government, such as at the municipal and local levels, the

departure from QBS requirements in favor of a cost-focused procurement process becomes even more noticeable.

A/E professionals have long been in disagreement with clients over the inclusion of cost proposals during the procurement process, and several professional associations have argued that procurement methods that focus on lowest-cost design options are more susceptible to shortsighted decisions that put quality secondary to the cheapest cost (Fleenor and Hall 2002; ACEC 2006). In instances where owners insist on evaluating cost, A/E professional associations have responded by establishing guidelines for how cost should be treated to ensure an effective, fair, and transparent selection process. The two most widespread recommendations are that owners should (1) limit the weight of pricing such that it does not dominate the selection outcome and (2) use a two-envelope system, in which cost is held separate from qualifications submissions during the evaluation stage, to ensure the evaluation process is not unduly biased towards lowest cost.

Little research has investigated the application of these guidelines within the industry. The objective of this study was to address this gap by analyzing the selection outcomes from a data set of 122 publicly procured A/E projects across North America (consisting of 804 individual consultant proposals), where the owners' evaluation process followed the guidelines of limited cost weighting and a two-envelope procedure. Three main research questions guided the research. First, the study sought to understand whether these procurement guidelines would result in selection outcomes that differed drastically from traditional QBS methods, which may indicate that the inclusion of cost has a disproportionate effect on procurement outcomes. Second, the study sought to determine which specific evaluation criteria (including cost, schedule, and assorted qualifications-based criteria) achieve the greatest level of differentiation between competing A/E proposals. A corresponding hypothesis posited that such evaluation criteria would achieve varying levels of differentiation. Finally, the study sought to analyze whether consultants who were had stronger qualifications also had higher price proposals. Individual consultant cost proposals were correlated with their associated evaluation scores on each qualifications based criterion, under the hypothesis that a directly proportional relationship would be found.

Findings contribute an understanding of the influence that cost evaluations have on selection outcomes. Related to the first research question, results revealed that selection results roughly mimicked QBS, such that the highest qualified consultant was selected in the majority of cases; however, a small minority of projects was awarded to consultants bidding the lowest who did not have the highest evaluated qualifications. Second, the relative effectiveness of evaluation criteria in differentiating between consultant proposals showed that cost, schedule, and interview criteria achieved the greatest level of differentiation, followed by technical proposals. This contributes an understanding of which evaluation criteria that A/E consultants can focus on to have the greatest chance of differentiating their value proposition to owner evaluation committees. Third, in response to owner-perceived concerns that QBS may result in higher upfront costs (Christodoulou et al. 2004), this study ultimately found no correlation between consultant cost submissions and owner evaluations of qualifications-based criteria. This finding contributes the understanding that owners should not assume that greater consultant qualifications automatically come at a greater cost.

LITERATURE REVIEW

The literature review is divided into two main areas. First, the perception of QBS held by various professional associations is reviewed. Second, the main research objectives and associated findings of previous studies are summarized to identify gaps in the current body of knowledge related to procurement.

The official position of professional associations within the A/E industry has been to promote the usage of QBS methods. The American Institute of Architects (AIA) strongly supports QBS for procuring professional design services for public projects, based on their reasoning that "QBS provides owners with a selection process that is not only straightforward and easy to implement, but is objective and fair" (AIA 2011). Similarly, the ACEC has highlighted QBS as the "cornerstone" of their procurement policies (ACEC 2006). The American Public Works Association (APWA) believes that procuring design services using qualifications rather than cost "fosters greater creativity and flexibility, and minimizes the potential for disputes and litigation" (APWA 2008). The National Society of Professional Engineers (NSPE) "strongly" supports the Brooks Act and recommends that the procurement of design professional services in all sectors should utilize QBS methods (NSPE 2016). The American Water Works Association (AWWA) specifically encourages water utilities to procure design services on the basis of QBS (AWWA 2016).

Studies on the topic of A/E procurement were found to be scarce. The current body of knowledge is limited to studies that focus on the areas of industry trends towards commoditization, multi-criteria decision-making models, prequalification procedures, and performance of A/E professionals during the design phase. The first focus area was commoditization of A/E professional services. Hampton (1994), Horns and Jenkins (2011), and Parks (2006) discussed how A/E professional's services are considered to be a commodity rather than a professional service. They further emphasized that professional services cannot be quantified; hence, A/E professionals should be compensated on value rather than time.

The second area focused on multi-criteria decision making for the selection of professional consultants. Cheung et al. (2002), Nguyen et al. (2008), and Ling (2003) developed models for procurement of A/E services using different evaluation criteria, whereas Sullivan et al. (2010) developed a best value (BV) Performance Information Procurement System (PIPS) for procuring A/E professionals.

The third focus area was prequalification procedures for A/E services. Ng and Chow (2004a) identified 28 preselection criteria that can be used for prequalifying A/E professionals. Feldmann et al. (2008) developed a regression model for predicting the bid costs of A/E professional services for various university projects.

Last, the fourth focus area was the performance of A/E professionals within various procurement systems and delivery methods. Ling (2002) and Ng and Chow (2004b) developed models for predicting A/E performance based on different project attributes of A/E professionals. Oyedele and Tham (2007) and Chow and Ng (2007) utilized a survey methodology to study A/E performance via qualitative performance metrics.

The most well-known and comprehensive research study on QBS was conducted by Chinowsky and Kingsley (2009). The study had a sample of 42 projects that analyzed and compared the benefits of QBS to non-QBS projects. The results showed that QBS projects had a construction cost growth and schedule growth of 3% and 8.7%, respectively. When owners were surveyed, 93% rated the success of their final project outcomes as “high” or “very high” in terms of project quality when QBS was used. Furthermore, QBS was identified as the preferred procurement method for high-risk projects. The other most relevant research, conducted by Christodoulou et al. (2004), analyzed 162 A/E projects from New York City that were procured using a two-envelope BV procurement method. Christodoulou et al. rejected the city’s claim that injecting cost competition to A/E selections had saved significant tax payer dollars. The city’s publicized data set was found to be skewed because it contained a substantial number of nonprofessional service projects. In addition, the city had not considered the final cost of project when calculating cost savings, which would realistically be valued at 1.67% of their claimed savings of $892 million.

Professional associations have strongly favored the procurement of professional design services using QBS because it incentivizes quality over cost. In addition, previous literature has studied the commoditization, multi-criteria decision-making procurement, prequalification, and performance of A/E professionals, but not much research was focused on the influence of cost within the BV procurement method, which created a gap in the literature.

RESEARCH OBJECTIVES AND HYPOTHESES

Point of Departure

The existing literature related to procurement of A/E professionals has been primarily limited to small sample sizes. The most prominent study of QBS, by Chinowsky and Kingsley (2009), analyzed 42 projects for their performance in cost, schedule, and quality. Among these projects, 33 were procured via QBS and 4 were procured using BV. For the BV projects, the study did not specify whether the recommended evaluation guidelines for A/E services were followed. Christodoulou et al. (2004) analyzed 162 projects for cost and schedule performance and ultimately determined that evaluation of consultant cost proposals did not result in significant project savings. Their study concluded that most projects in the sample contained nonprofessional services (approximately 73%) and the final cost of the project was ignored when calculating the savings. Other studies of A/E procurement mainly studied multicriteria decision-making models. The present study aims to address this gap by analyzing a relatively larger sample size, with 102 BV projects (plus an additional 20 QBS projects), which followed the recommended guidelines of limiting cost evaluations to less than 30% of the total weight and utilizing a two-envelope evaluation process.

Research Question 1

A/E professionals are generally opposed to a substantial introduction of cost within owner procurement methods. Christodoulou et al. (2004) stated that when low bidding is used, the owner essentially commoditizes A/E professionals. In some cases, when cost was used as a dominant evaluation factor, the effect was that consultants were essentially being selected on the basis of lowest bidder (Chinowsky and

Kingsley 2009). The cost factor has crept into owner procurement processes to the extent that El Wardani et al. (2006) assumed projects to be QBS if the cost weight was as high as 50% of the total evaluation. Internationally, in order to mediate the low bid selection criteria of consultants, the World Bank states that if A/E professionals are chosen based on a combination of cost and qualifications, the weighting of cost should not exceed 30% (Fleenor and Hall 2002). Yet little research has documented the outcomes when owners operate within those guidelines. Therefore, it becomes important to analyze the effect of cost on the procurement process, even after limiting the weight to 30% or less. This led to the development of the first research question:

What are the selection outcomes of the two-envelope BV procurement processes for architectural and engineering services; furthermore, does the inclusion of cost as an evaluation criterion have a disproportionate effect on selection outcomes, such that the owners trend towards the selection of lowest bid (LB) or non-highest qualifications?

Research Question 2

The purpose of using the two-envelope evaluation procedure was to select the consultant that presented the best value combination of cost and qualifications. With 70% or more of the evaluation hinging upon qualifications-based criteria, it is important for the owner to choose criteria that effectively differentiates the level of qualifications between competing consultants. Therefore, the present study sought to determine which evaluation criteria achieved the greatest differentiation in owner evaluation scores among competing consultants. This led to the development of the second research question and its corresponding hypothesis statement:

Which evaluation criteria achieve the greatest level of differentiation between competing A/E proposals?

Hypothesis Statement 1 H1: Individual evaluation criteria will result in various levels of differentiation between competing A/E consultants.

Research Question 3

Owners often assume that higher bid costs are encountered when a consultant demonstrates greater qualifications than competing consultants. This assumption can cause uneasiness among owners in QBS, where the owner is unable to solicit pricing information other than from the selected consultant.

Does the selection of greater qualifications among A/E consultants correspond with higher cost proposals?

Hypothesis Statement 2 H2: A directly proportional relationship exists between the cost criteria and other qualifications criteria within individual consultant proposals.

METHODOLOGY

Data Collection

The data sample was collected from various A/E projects across North America at the state, city, and municipal levels as well as institutions of higher education. The sample size consisted of 122A/E projects, of which 102 were procured via a two-envelope BV method and the remaining 20 were procured using QBS.

All projects in the sample used virtually identical evaluation criteria with similar weighting schemes. The main evaluation criteria included cost, proposed schedule, a written technical proposal, past performance of the consultant and their project team, consultant team interviews, and related experience (RE) submissions of the consultant firm's previous projects. Within the BV-procured projects, cost was limited at 10–30% of the total evaluation weight and a two-envelope evaluation process was used. Given that this is in accordance with recommendations of A/E professional associations, the data sample was deemed to be an important contribution to the literature.

Definition of Variables

Evaluation Criteria

This research focused on six common evaluation criteria of cost, proposed schedule, technical proposals, past performance, interviews, and related experience.

Cost. The cost criterion was the total design cost of the project that the consultant quoted. Cost was evaluated on the basis of LB, where the LB received the highest evaluation score, while the remaining consultants received score deductions in inverse proportion compared with the LB. For purposes of data analysis, cost was normalized on a per-project basis as the percentage of the LB and percentage relative to the average bid. Cost as percentage of LB (% LB) for each project was calculated using the following formula:

Cost (% LB) = (Lowest bid cost of the project– proposed cost of consultant) / (Lowest bid cost of the project). **(1)**

Similarly, cost as percentage of average bid [percentage average (% Avg.)] for each project was calculated using the following formula:

Cost (% Avg.) = (Average bid of the project- proposed cost of consultant)/
(Average bid of the project) **(2)**

Schedule. This criterion consisted of an anticipated project duration from each proposing consultant. Similar to cost, the consultant with the minimum proposed schedule was assigned the highest evaluation score, whereas all other consultants were scored using the inverse proportion method. Schedule evaluations were normalized to a percentage beyond of the minimum proposed schedule duration on a

per-project basis. Schedule normalized (Norm.) for each project was calculated using the following formula:

Schedule (Norm.) = (Least days for the project/ proposed days of the consultant) **(3)**

Technical Proposal. Technical proposals (TPs) were typically composed of a narrative of each proposing consultant's execution plan, discussion of project-specific risk mitigation strategies, and identification of potential scope alternatives and innovative approaches to the project.

Past Performance. Past performance was a reflection of the consultant's previous experience with similar projects and clients. Within this study, past performance was evaluated as a survey of each consultant's previous clients as to their satisfaction.

Interviews. Interviews were conducted with key personnel from each shortlisted consultant's proposed project team. Interviews were mostly conducted with the top three to five consultants per project.

Related Experience. RE required consultants to submit a summary of their most similar reference projects. Miscellaneous Criteria. At times, owners utilized other evaluation criteria. Some criteria were dictated by policy requirements, such as Women and Minority Business Enterprise (WMBE), consideration. However, these criteria were not considered within the analysis because they were typically assigned a relatively low weight (5– 10%) and were inconsistently applied across the data set.

Coefficient of Variation

For the second research question, the coefficient of variation (CV) was employed to determine which of the evaluation criteria achieve the greatest differentiation. CV is a measure of statistical dispersion that describes the standard deviation relative to the mean. To elaborate further, the CV for each evaluation criteria tells which criteria give the widest dispersion of scores for a project on a normalized basis. Because the resulting value is a dimensionless quantity, it can easily be compared with all criteria irrespective of the specific unit in which they were measured.

Evaluation Procedures Used in the Data Sample

The projects within the data set all followed a virtually identical two-envelope BV evaluation process. With the release of a request for proposal (RFP), a presubmittal conference was held for potential proposing consultants, where the scope of work, evaluation criteria, and the evaluation process for the RFP was discussed. On the closing date, bids were received in two-envelopes. The first envelope contained all qualifications documents (technical proposals, past performance, and RE) and schedule days, whereas the second contained the consultant's cost proposal. Once received, technical proposals and RE documents were sent out to the evaluation committee for scoring. Technical proposals were kept anonymous from the evaluation committee for unbiased and transparent evaluations. The evaluation committee was also told to evaluate the proposals independently so that they were not influenced by any other evaluator's decision.

In accordance with the two-envelope evaluation process, the procurement officer separately inserted the evaluation scores for cost, schedule, and past performance into an evaluation matrix. After the scores for technical proposal and RE were received, they were inserted in the file, and the top three to five rated consultants were shortlisted to participate in the interview process. Key personnel from each of the shortlisted consultant's project team were interviewed by the owner's evaluation committee.

After the interview scores were inserted into the matrix, a linear relationship model (LRM) was used to calculate the raw evaluator provided scores into the weights assigned for each criterion. For each project, the consultant who earned the greatest amount of evaluation weights was selected and awarded the contract.

Data Characteristics

Table 1 gives the distribution of evaluation weights for each evaluation criteria as published in the owner solicitation documents. Interviews and technical proposals were the top two highest weighted criteria with mean weights of 33.86% and 33.48%, respectively. The mean value for cost weights was 19.58%, which is well within industry guideline of using 30% as a maximum. Past performance criteria had a mean weight of 19.9%, schedule was assigned a mean weight of 13.69%, and RE was assigned the least weight in the evaluation process with a mean value of 9.55%.

Table 1. Descriptive analysis of evaluation criteria weights

Evaluation criteria	*n*	Minimum (%)	Maximum (%)	Mean (%)	Median (%)	Standard deviation (%)
Cost	102	0[a,b]	40.00	19.58	20.00	6.80
Schedule	82	5.00	30.00	13.96	14.64	5.56
Technical proposal	122	15.00	55.00	33.48	35.00	6.59
Past performance	121	5.00	50.00	19.90	20.00	9.14
Interview	54	20.00	55.00	34.86	35.00	7.82
Related experience	65	5.00	26.67	9.55	6.67	5.61

[a]0 refers to projects procured via pure QBS.
[b]For BV, the minimum weight was 10%.

METHOD OF ANALYSIS

Descriptive Analysis of BV Procurement Outcomes

Descriptive statistics were used to analyze 122A/E projects to investigate the first research question. Frequency analysis was conducted for the characteristics of the selected consultants, including the frequency selections of consultants who were the LB, best qualifications (BQ), and overall best score (BS). It consisted of frequency in percentage of selected LB, BQ, and BS for BV and QBS projects. LB referred to selected consultants who had bid the lowest cost among all competing consultants on a per-project basis. Similarly, BQ referred to consultants who were best in qualification and were selected. Lastly, BS consultants were those who had the highest overall evaluation score considering the combination of both cost and qualifications-based criteria.

The projects were also analyzed to determine the selected consultants' ranking for each evaluation criteria compared with competing firms (SR), selected consultants' score (SS), average project score (AS), and average LB score (LS) for the six evaluation criteria. The average of these values is summarized in Table 2 with two more quality measures, the "differential from the average bid" (DAB) and the "differential from the LB" (DLB). The DAB and DLB described the added value a selected consultant brought to the project when compared with the average bid and LB. The DAB and DLB were calculated using the following formulas:

DAB = Average of (((SS-AS)/AS) x 100) **(4)**

And,

DLB = Average of (((SS-LS)/LS) x 100) **(5)**

Table 2. Differential between selected consultants and competing proposals

Evaluation criteria	Average selected consultants' ranking	Average selected consultants' score	Average of average scores	Average of lowest bid scores	Differential from average bid (%)	Differential from lowest bid (%)
Cost	2.21	—	—	—	−15.74	44.93
Schedule	2.46	—	—	—	−15.79	−5.11
Technical proposal	1.93	67.97	59.89	59.63	14.42	34.25
Past performance	2.84	90.51	88.12	86.10	3.07	15.15
Interview	1.24	84.47	66.92	50.64	30.62	180.07[a]
Related experience	1.20	90.66	90.54	89.64	0.52	3.73

[a] $n = 17$ projects in which lowest bid was shortlisted for the interview and was not the selected consultant.

Differentiation in Scores for Evaluation Criteria Using the Kruskal-Wallis H-Test

The Kruskal-Wallis H-test, which is also known as the one-way ANOVA on ranks test, is a rank-based nonparametric test. The test was utilized to answer the second research question and first hypothesis statement by determining whether differentiation in scores between the six evaluation criteria was statistically significant. Pairwise comparison was adopted as a post hoc test for the Kruskal-Wallis H-test. Advantages of the Kruskal-Wallis H-test are that it does not assume the data to be normally distributed and also that it is not substantially affected by outliers. Moreover, the independent variables are assumed to be categorical, whereas the dependent variable should be continuous in nature.

Relationship among Evaluation Criteria Using Spearman's Correlation

Spearman's correlation coefficient, also known as Spearman's rank-order correlation, measured the strength and direction of the association between two continuous or ordinal variables. Spearman's correlation was used to answer the last research question and the second hypothesis question, by analyzing the relationship of cost with other qualifications criteria. Like the Kruskal-Wallis H-test, Spearman's correlation does not assume data to be normally distributed and is not substantially affected by outliers. The test does assume for the two variables to have a monotonic relationship.

RESULTS

Selection Outcomes in Best Value Procurement of A/E Consultants

Selected consultants were ranked first in two of the six evaluation criteria and submitted better cost and schedule proposals compared with average bids. In almost a quarter of the projects, the selected consultants were simultaneously ranked as the LB and the BQ. In Table 2, it can be seen that selected consultants were placed, on average, first in interview and RE; second in cost, schedule, and technical proposals; and third in past performance. In the BV procurement system, selected consultants always resulted in favorable differentials when compared with the average bid or the LB. It can also be seen in Table 2 that in comparison with the average bids' evaluation scores, selected consultants were 15.74% less costly, proposed a 15.79% faster schedule, and were 14.42, 3.07, 30.62, and 0.52% more qualified in technical proposals, past performance, interviews, and RE, respectively. Similarly, when compared with the LB's scores for the six evaluation criteria, selected consultants were 44.93% more costly; had a 5.11% faster schedule; and were 34.25, 15.15, 180.07, and 3.73% more qualified in technical proposals, past performance, interviews, and RE, respectively.

Attaining first rank in the interview and RE criterion resulted in a very high probability of being selected. Fig. 1 provides the rank frequency distribution of selected consultants per evaluation criteria in the six evaluation criteria. The selected consultant was ranked first in interviews 81% of the time compared with 13% and 6% for second and third, respectively. No consultants were selected if they were ranked fourth or lower. The consultant ranked first in RE was selected 88% of the time, while the probability of selection decreased drastically if ranked lower.

In a BV procurement process, the BQ consultant was selected almost half the time. Table 3 provides the arrangement of the selected consultants by rank in cost and qualifications criteria. Within the table, cost refers to the ranking of the cost proposal submitted within the selected consultant's proposal. The qualifications criteria items were taken as a combination of cumulative evaluation scores for technical proposals, interview, past performance, and RE. Across the data set, the BQ consultant was selected 54% of the time, whereas the LB was selected 48% of the time. In 22% of cases, the selected consultant was both the LB and the BQ. It was also noticed that the best evaluation score consultant was awarded the contract only 84.31% of the time. In the remaining 15.69% of cases, the owners were incentivized to select the second ranked consultant (or third ranked in just one case) based on the justification that the BS consultant was either too high (or even too low) in their cost proposal relative to competing consultants.

Evaluation Criteria Effectiveness in Differentiating Consultant Qualifications

The CV was used as a measure of dispersion among competing consultant proposal elements within the data sample. This measure suited the need of the analysis to determine which evaluation criteria resulted in the greatest differentiation in evaluation scores between competing consultant proposals. Table 4 gives the descriptive statistics for the CV values for the six evaluation criteria. Due to the fact that the data set was not normally distributed and nonparametric statistical tests were conducted, the median is reported as the measure of central tendency rather than the mean. Among all evaluation criteria, schedule had the highest median CV value of 23.79% (mean of 26.86%), followed by cost and interview with 22.50% (mean

of 25.03%) and 17.10% (mean of 24.13%), respectively. Technical proposals achieved a median CV value of 14.95% (mean of 17.14%). Past performance achieved a CV value of 4.80% (mean of 8.23%), and RE had a median value of 0% (mean of 7.62%).

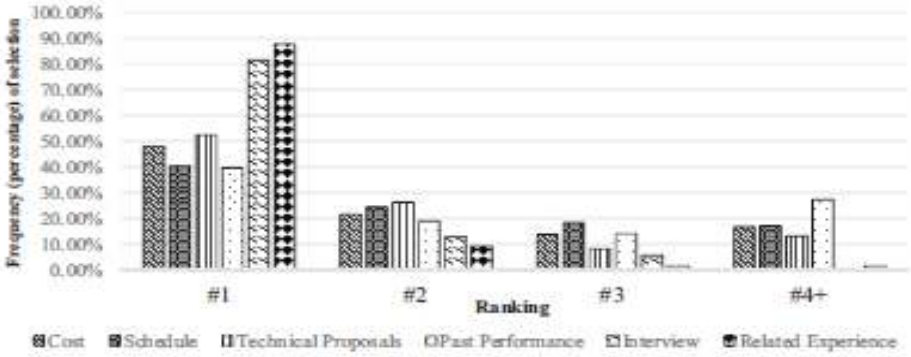

Fig. 1. Ranking of selected consultants for all evaluation criteria.

Table 3. Cost-qualification ranking matrix for selected consultants

Selected consultant evaluation ranking (%)	Number 1 qualification (%)	Number 2 qualification (%)	Number 3 qualification (%)	Number ≥4 qualification (%)	LB consultant (%)
Number 1 cost	21.57	17.65	4.90	3.92	48.04
Number 2 cost	8.82	7.84	0.98	3.92	
Number 3 cost	10.78	0.00	0.98	1.96	
Number ≥4 cost	12.75	2.94	0.00	0.98	
BQ consultant	53.92				

Note: Best evaluation score was selected 84.31% of the time. LB = lowest bid; and BQ = best qualification.

Table 4. Descriptive statistics for coefficient of variation

Evaluation criteria	*n*	Mean (%)	Median (%)	Standard deviation (%)	Minimum (%)	Maximum (%)	Measure
Cost	102	25.03	22.50	15.31	0.00	78.50	USD
Schedule	82	26.86	23.79	17.50	0.00	98.37	Days
Technical proposal	122	17.14	14.95	11.56	0.00	68.60	0–100
Past performance	121	8.23	4.80	8.91	0.00	57.80	0–100
Interview	54	24.13	17.10	20.76	0.00	90.73	0–100
Related experience	65	7.62	0.00	15.87	0.00	70.70	0–100

Table 5. Kruskal-Wallis *H*-test result for evaluation criteria differentiation

Evaluation criteria	Sample size (*n*)	Median CV (%)	χ^2-Value	Degrees of freedom	*p*-Value
Cost	102	22.50	173.666	5	0.000[a]
Schedule	82	23.79			
Technical proposal	122	14.95			
Past performance	121	4.80			
Interview	54	17.10			
Related experience	65	0.00			

[a]Statistically significant at the 0.01 level.

The Kruskal-Wallis H-test was used to determine if there were statistically significant differences in median CV values between the six evaluation criteria. The distribution of CV values was found to be similar for the six evaluation criteria, as assessed by a visual inspection using a box plot. As presented in Table 5, median

CV values for the six evaluation criteria were statistically significant between groups with x2 (5) = 173.666 and p = 0.00.

Post hoc tests were performed to further investigate the statistically significant differences in CV between each pair of evaluation criteria. Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Post hoc analysis, as shown in Table 6, revealed statistically significant differences in median CV values for $p<0.05$. Based on the analysis, the three evaluation criteria of cost, schedule, and interviews achieved the greatest differentiation in scores, with median CV values of 22.50, 23.79, and 17.10%, respectively. Next, technical proposals achieved moderate differentiation, with a median CV value of 14.95%. Past performance and RE with median CV values of 4.80% and 0.00%, respectively, gave the least amount of differentiation in evaluation scores.

Table 6. Post hoc test results for evaluation criteria differentiation

Pairwise comparison of evaluation criteria CV central tendencies	*p*-Value
Related experience versus past performance	0.272
Related experience versus technical proposal	0.000[b]
Related experience versus interview	0.000[b]
Related experience versus cost	0.000[b]
Related experience versus schedule	0.000[b]
Past performance versus technical proposal	0.000[b]
Past performance versus interview	0.000[b]
Past performance versus cost	0.000[b]
Past performance versus schedule	0.000[b]
Technical proposal versus interview	1.000
Technical proposal versus cost	0.015[a]
Technical proposal versus schedule	0.007[b]
Interview versus cost	1.000
Interview versus schedule	1.000
Cost versus schedule	1.000

[a]Statistically significant at the 0.05 level.
[b]Statistically significant at the 0.01 level.

Relationship between Consultant Qualifications and Bid Cost

Cost was found to have no practically significant correlation with any qualifications-based evaluation criteria. Spearman's correlation coefficient was used to determine if a relationship existed between each consultant's cost proposal and their evaluation scores received on qualifications-based criteria. Table 7 presents the Spearman's correlation coefficient for each combination of evaluation criteria. Cost (% Avg.) showed a slight negative correlation with schedule (Norm.) with Spearman's correlation (rho) rs = –0.177, which was significant at 0.01. Similarly, cost (% LB) was also found to be statistically significant to schedule (Norm.) and past performance with rs = –0.287 and –0.117, respectively, both significant at 0.01. Despite their statistical significance, the extremely weak correlation coefficients indicate that the associations are of no practical impact for industry practitioners. No other qualifications criteria showed significant result with cost.

Table 7. Relationship between consultant qualifications and bid cost using Spearman's correlation

Number	Evaluation criteria	1	2	3	4	5	6	7
1	Cost (% Avg.)	1.000	—	—	—	—	—	—
2	Cost (% LB)	—	1.000	—	—	—	—	—
3	Schedule (Norm.)	−0.177[b]	−0.287[b]	1.000	—	—	—	—
4	Technical proposal	0.061	−0.015	0.088	1.000	—	—	—
5	Past performance	0.029	−0.117[b]	0.113[a]	0.095[a]	1.000	—	—
6	Interview	0.075	0.165	0.006	0.426[b]	0.056	1.000	—
7	Related experience	−0.002	0.026	0.079	0.102[a]	−0.180	0.296	1.000

[a]Correlation is significant at the 0.05 level (2-tailed).
[b]Correlation is significant at the 0.01 level (2-tailed).

DISCUSSION

Selection Outcomes in Best Value Procurement of A/E Consultants

BV projects mimic the results of QBS almost half the time, as the BQ consultant was selected almost half the time, compared with every time in QBS. This indicates that even when cost is used as an evaluation criterion, owners today generally want better-qualified consultants who can understand the scope of work, refine it, look for every possible risk involved, and add as much value to the project as possible.

Similarly, LB consultants were awarded the contract almost 48% of the time. Of this 48%, almost 21.5% of the time the LB was also the BQ; therefore, only 26.5% of the consultants were the LB and not the BQ.

Based on the descriptive results, it can be concluded that the inclusion of cost as an evaluation criterion does have disproportionate effect on selection outcomes, such that selections are slightly skewed towards the LB or non-highest qualified consultant. This can be seen from the fact that the LB consultant was selected in 26.5% of cases, which exceeds the mean cost weights of 20%. Further, the best-qualified consultant was selected in only 54% of cases compared with the mean weight of 75% assigned to

Evaluation Criteria Effectiveness in Differentiating Consultant Qualifications

Cost, schedule, and interview achieved the greatest differentiation in evaluation scores. Because every consultant proposed bids based on their understanding of the owner's design intent, and perhaps because the owners rarely published their project schedule and budget constraints in the RFP, a higher differentiation existed for cost proposal and schedule proposal among all the proposing consultants. Furthermore, interviews were generally designed in a manner where a detailed review of the submitted technical proposals was discussed. Therefore, the evaluation committee scored consultants based on the specifics of the project, resulting in high scores for highly qualified consultants.

Technical proposals achieved a moderate differentiation in evaluation scores, whereas past performance and RE achieved the least differentiation in score. Because past performance was scored by past clients for consultants' performance in their project, consultants only submitted surveys that scored them highly, resulting in less differentiation. Moreover, a majority of the consultants in a project got the highest RE

score, which gave an almost zero standard deviation, resulting in zero or very low CV values researchers rejected the null hypothesis and therefore accepted H11 such that different evaluation criteria do result in various levels of differentiation between competing A/E consultants.

Relationship between Consultant Qualifications and Bid Cost

Higher cost proposals were not directly associated with qualification based evaluation scores. Cost was compared with other qualifications criteria using two cost measures, cost (% Avg.) and cost (% LB). Although several correlations were found to be statistically significant with these measures, the extremely low values of the correlation coefficients essentially rendered these relationships to be of no practical application to the industry. The researchers failed to reject the null hypothesis, and therefore H20 was retained, and it was concluded that no direct relationship existed between cost proposals and other qualifications-based criteria.

CONCLUSIONS

As the inclusion of cost as an evaluation criterion continues to be a controversial topic in within the design industry, it has become important to determine extent to which the inclusion of cost incorporation affects the procurement process. Little research has documented selection outcomes of BV procurement methods in the context of A/E professional services, particularly within the bounds of industry guidelines that recommend a two-envelope evaluation process with cost proposals limited to a maximum of 30% weighting. This research fills in the gap by analyzing a relatively larger sample size and geographic location and analyzing projects that used industry BV guidelines.

A total of 122A/E projects procured using BV and QBS methods were analyzed to determine characteristics of the selected consultants using descriptive analysis, frequency distribution, matrix, and measure of central tendency. Kruskal-Wallis H-tests were used to find the level of differentiation achieved by different evaluation criteria. Last, Spearman's correlation coefficient was used to determine whether cost criterion relationships exist with any other qualifications criteria.

In BV projects, more than half the time, the BQ was selected, compared with 100% in a QBS system. Also, 48% of the time, the LB was selected. However, the percentage of consultants selected purely on cost was 27%. Therefore, it can be concluded that the inclusion of cost as an evaluation criterion does have disproportionate effect on selection outcomes. When determining the level of differentiation, cost proposals, schedule proposals, and interviews attained the greatest differentiation, whereas technical proposals showed a moderate differentiation. Past performance and RE resulted in the least differentiation. The researchers rejected the null hypothesis, and therefore H11 was accepted. Cost was found to have no relationship with any of the qualifications criteria. Therefore, the study contradicts the perception that greater qualifications correspond with higher design costs. The researchers failed to reject the null hypothesis, and therefore H20 was retained.

CONTRIBUTIONS

Contributions to the Body of Knowledge

Previous research has not assessed the outcomes of BV procurement for A/E consultants, particularly based upon industry recommended guidelines of limited weighting towards cost proposals and a two-envelope evaluation process. This study analyzed a sample size of 122A/E professional projects across North America, which was relatively larger than previous studies. The unit of measure employed was novel where the evaluation committee scores of all 804 competing consultant proposals were analyzed for each project.

Contributions to Industry Practitioners

The outcomes of BV procurement for A/E services roughly mirrored the results of traditional QBS. Yet in order to achieve these outcomes, owners must be careful to structure their BV procurement methods within recommended industry guidelines for A/E selections. Results also showed that interviews and technical proposals achieved higher differentiation in scores than other qualifications-based criteria; therefore, owners are recommended to focus their evaluation of A/E qualifications on the quality of the project team individuals that would be assigned to their project along with the technical proposal of the project-specific means, methods, and innovations the team would employ.

Limitations and Recommendations for Future Research

There are several limitations to this study. First, this study only focuses on a two-envelope BV procurement process, and therefore the results do not mimic the trends of different variants of BV procurement that may exist in practice. Second, project budget and schedule information was not always provided by the owner, which hampered the researchers' ability to determine how the submitted cost and schedule proposals related to the owner's anticipated budget or schedule.

Furthermore, this study also opens different avenues of research, as future researchers are recommended to focus on detailed analysis of how the owner-provided budget, cost, and schedule affect the submitted cost and schedule proposals. Different project characteristics such as level of owner-provided design, delivery method, and contract type may also be investigated. Key personnel interview scores can be investigated in detail to determine the level of differentiation (measured in terms of CV) achieved by interviewing different individual personnel within the design team.

LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Despite the above stated conclusions and contributions, there are a few limitations to this research. This study only focuses on two-envelope BV procurement process, and therefore the results do not mimics the trends of different variants of A/E BV procurement. Also, project budget and schedule information was not always provided by the owner, which hampered in determining if the differentiation in cost and schedule proposal, was below, on or above the owner's budget or anticipated schedule.

Furthermore, this study also opens different avenues of research, as future researchers can focus on detailed analysis of how the owner-provided budget, cost, and schedule affect the submitted cost and schedule proposals. Different project characteristics such as level of design, delivery method, etc. can also be investigated for differentiation achieved in scores. Key personnel interview scores can be investigated in detail to determine the level of differentiation achieved by different personnel.

REFERENCES

ACEC (American Council of Engineering Companies) (2006). ACEC Policy Statement Selection of Design Professionals on the Basis of Qualification. ACEC, Washington, DC.

AIA (The American Institute of Architects) (2011). Issue Brief Qualification Based Selection. AIA, Washington, DC.

APWA (American Public Work Association) (2008). Qualifications Based Selection of Professional Services Consultants. APWA, Washington, DC.

AWWA (American Water Works Association). (2016). "Qualifications-Based Selection of Professional Services." Policy Statement, < http://www.awwa.org/about-us/policy- statements/policy-statement/articleid/4039/qualifications-based-selection-of-professional- services.aspx> (July. 29, 2016).

Cheung, F. K. T., Kuen, J. L. F., and Skitmore, M. (2002). "Multi-criteria evaluation model for the selection of architectural consultants." Construction Management and Economics, 20(7), 569-580.

Chinowsky, P. S., and Kingsley, G. A. (2009). "An analysis of issues pertaining to qualifications- based selection." American Council of Engineering Companies and American Public Works Association, Washington, DC.

Christodoulou, S., Griffis, F., Barrett, L., and Okungbowa, M. (2004). "Qualifications-Based Selection of Professional A/E Services." J. Manage. Eng., 10.1061/(ASCE)0742- 597X(2004)20:2(34), 34-41.

Chow, L, K., and Ng, S, T. (2007). "Expectation of performance levels pertinent to consultant performance evaluation." International Journal of Project Management, 25(1), 90-103.

Feldmann, M., Chrusciel, D., Pohlmann, A., Shelley, M., II, McCool, K., Morton, A., and Ahoy, C. (2008). "Architectural and Engineering Fees from the Public Institutional Perspective." J. Manage. Eng., 10.1061/(ASCE)0742-597X(2008)24:1(2), 2-11.

Fleenor, C., and Hall, S. (2002). "The state of QBS: battles at the local and national level are challenging how firms do business." Engineering Inc., 13(5), 16–19.

Hampton, D. (1994). "Procurement Issues." J. Manage. Eng., 10.1061/(ASCE)9742- 597X(1994)10:6(45), 45-49.

Horns, K., and Jenkins, R. (2011). "Is the Profession of Civil Engineering Becoming a Commodity? You Should Know the Answer." Leadership Manage. Eng., 10.1061/(ASCE)LM.1943 5630.0000089, 40-44.

Ling, Y, Y. (2002). "Model for Predicting Performance of Architects and Engineers." J. Constr. Eng. Manage., 10.1061/(ASCE)0733-9364(2002)128:5(446), 446-455.

Ling, Y, Y. (2003). “A conceptual model for selection of architects by project managers in Singapore.” International Journal of Project Management, 21(2), 135-144.

Nguyen, T. H., and Shehab, T. (2008). “Selecting an architecture-engineering team by using fuzzy set theory.” Engineering, Construction and Architectural Management, 15(3), 282-298.

NSPE (National Society of Professional Engineers). (2016). “Qualifications-Based Selection of Engineering Services.” Issues and Advocacy, < https://ww.nspe.org/resources/issues and-advocacy/action-issues/qualifications-based-selection-engineering-services> (July. 29, 2016).

Oyedele, L, O., and Tham, K, W. (2007). “Clients’ assessment of architects’ performance in building delivery process: Evidence from Nigeria.” Building and Environment, 42(5), 2090-2099.

Parks, G. (2006). "At a Crossroads— Is Value-based Compensation the Answer?" Leadership Manage. Eng., 10.1061/(ASCE)1532-6748(2006)6:4(144), 144-149.

Sullivan, K, T., Savicky, J., and Carey, B. (2010). “Best-Value Process Implementation at the City of Peoria: Five Years of Research Testing.” Journal for the Advancement of Performance Information and Value, 2(1), 23-32.

El Wardani, M., Messner, J., and Horman, M. (2006). "Comparing Procurement Methods for Design-Build Projects." J. Constr. Eng. Manage., 10.1061/(ASCE)0733- 9364(2006)132:3(230), 230-238.

Thomas Ng, S., and Chow, L. (2004a). "Framework for Evaluating the Performance of Engineering Consultants." J. Prof. Issues Eng. Educ. Pract., 10.1061/(ASCE)1052- 3928(2004)130:4(280), 280-288.

Thomas Ng, S., and Chow, L. (2004b). "Evaluating engineering consultants’ general capabilities during the pre-selection process - a Hong Kong study." Engineering, Construction and Architectural Management, 11(3), 150-158.