Contractor selection using BVP in the construction industry
Case studies at the Dutch Ministry of Infrastructure

Jeroen van de Rijt
Scenter Management Consultants
Email: rijt@scenter.nl

Wiebe Witteveen
Dutch Ministry of Infrastructure
Email: wiebe.witteveen@rws.nl

Abstract

Public Contracting Authorities (PCA's) in the Netherlands are struggling to find a successful way to select the right supplier for infrastructural projects. The paper argues that PCA's lack information to exactly quantify their real preferences. This paper aims at researching the question whether PCA's can use a different way of applying MEAT, without an ex-ante decomposition of all MEAT criteria. This paper involves six case studies at The Dutch Ministry of Infrastructure using the methodology of Best Value Procurement /Performance Information Procurement System (BVP/PIPS). It concludes that not ex-ante knowing the utility curve can be overcome.

Keywords: Best Value Procurement, MEAT, utility curve, The Netherlands

Introduction

The European legislation demands from PCA's that tenders are transparent, objective and non-discriminatory. Within the European framework PCA's have a certain degree of freedom to select suppliers. In the past, procurement strategies based on the award criterion lowest price and detailed technical specifications have often been used to tender public works contracts. In many cases the outcome of these tenders has not been satisfactory. Choosing to tender with detailed specifications made low contractors' performance possible. Using the lowest price as an award criterion even led to fraudulent behaviour by contractors (Van Leeuwen, 2011). This is why PCA's have made a shift towards using Most Economically Advantageous Tender (MEAT) to award tenders in the last years. Rijkswaterstaat, part of the Dutch Ministry of Infrastructure and the largest PCA in the infrastructure market, has the ambition to become the leading agency within the Dutch central government. One of the means to reach this ambition is to award 90% of all infrastructure contracts on MEAT in 2012. Table 1 shows the progress since Rijkswaterstaat started in 2005 (Van Leeuwen, 2011)

	2006	2007	2008	2009
Total number of tenders	170	146	85	103
Tenders with MEAT	37 (22%)	41 (28%)	36 (42%)	49 (48%)
MEAT was lowest price	59%	66%	72%	65%

Table 1: MEAT at the Ministry of Transport

Using the methodology of MEAT certainly gives a client more opportunities than solely awarding on price, but applying MEAT does not always lead to a successful project (e.g. Palaneeswaran and Kumaraswamy, 2000).

1

When contracting authorities choose to award the contract to the MEAT they should determine economic and quality criteria, which as a whole determine which tender offers most value for money. The quality criteria require a form of performance measurement. A performance measurement is an ordered set of plausible performance levels set on a quantatitive or qualitative scale (Mateus et al., 2010). When using evaluation or MEAT-criteria Rijkwaterstaat uses both types of performance measurement:

- Quantative scale, such as traffic congestion in traffic loss hours or delivery time in days.
- Qualitative scale: such as table 2 which is frequently used in tenders by Rijkswaterstaat.

In defining evaluation criteria the following rules can be identified:
1. Alignment with project goals.
2. Organised in a hierarchal structure.
3. If necessary desegregation: evaluation criteria should be desegregated into subcriteria to further clarify the meaning of the criteria.
4. Intrinsic coherence of the evaluation model, such as isolability, non-redundancy and measurability.
5. The criteria should only represent project objectives over which tenderers have direct influence such as traffic congestion during building

Based on these five rules the conclusion appears to be that the quantitative scales are preferred over qualitative scales and the qualitative scales should be desegregated and be transformed into a scoring function that exactly represents the preferences of the PCA.

If qualitative scales are used, a common way of using MEAT is to decompose all MEAT criteria in detail prior to the bids. This notion assumes that the scoring rules before the tender are known and that the tender evaluation model is consistent with the real preferences of a PCA (Dreschler; 2009 and Mateus et al. 2009). This assumes that the PCA is aware of all his preferences (and the exact utility function) before knowing what the tenderers can offer. It has been argued that qualitative scales are subjective and ill defined, because a performance measure should be as clear and objective as possible. The performance measurement should be measurable and the contract documents should clearly and objectively define the aspects that are in competition (Mateus et al. 2010).

To be able to compare the performance levels with each other scoring functions will have to be created. Scoring functions allow the transformation of the performance levels set on each performance into partial scores set on an numeric scale representing attractiveness ("preference" or "value") between each performance level on a given criterion (Belton and Stewart, 2001). Scoring functions can be relative or absolute. Relative scoring functions define the score of a tender by comparing its performance with the ones of other tenders. This implies that it is impossible to score a tender without having knowledge of other tenders. The accountability for the relative attractiveness of a performance level is shifted from the PCA to the tenderer.

Table 2 shows an example of a qualitative scale with an absolute scoring function:

Level	Description
1	Only generic tasks are presented; without clear durations
2	Onlys generic tasks are presented; durations are clearly assigned and; an overall completion time of no more than 12 months is proposed
3	Most tasks are presented, divided by types of work; durations are clearly assigned

2

	and consistent with allocated resources; principal task relationships are presented and; an overall completion time of no more than 10 months is proposed.
4	Almost all tasks are presented, divided by types of work; durations are clearly assigned and consistent with allocated resources; all task relationships are presented and; an overall completion time of no more than 9 months is proposed.

Table 2 Example of a qualitative performance measure for a working plan quality criterion (Mateus, et al, 2010)

The advantage of the quality criterion is that it complies with the rules for defining evaluation criteria; the disadvantage is that this type of criterion leads to less differentiation, because all tenderers will adjust their bid for level 4. The consequence is that the award eventually is made on the lowest price, with the negative consequences sketched above.

This paper aims at researching the question whether PCA's can use a different way of applying MEAT, without an ex-ante decomposition of all MEAT criteria. This paper involves 6 case studies at The Dutch Ministry of Infrastructure in The Netherlands, with a total spend of circa $ 800 mln, using the methodology of Best Value Procurement /Performance Information Procurement System (BVP/PIPS).
The paper addresses the main differences between BVP/PIPS and more commonly used ways by The Dutch Ministry of Infrastructure to identify the MEAT. The paper describes the results of the 6 procurement processes. The structure of the paper is as follows: in the next section, the BPV/PIPS philosophy is laid out. Next, the case studies are described. This is followed by a description of the results of the case studies. The paper ends with conclusions.

Best Value Procurement
Best Value Procurement/Performance Information Procurement System (BVP/PIPS) was developed and refined by Dean Kashiwagi (Kashiwagi, 2011). Best Value Procurement is a process where both price and performance are considered instead of just price (and is in this way comparable with using MEAT). BVP/PIPS has been tested over 700 times, delivering construction services worth over $2.3B (1994-2010.) The results of the BVP/PIPS tests have been (Kashiwagi, 2009):

- 98% client satisfaction and no vendor caused cost deviation.
- Minimized up to 90% of the client's risk and project management.
- Vendors increased profits up to 100% without increasing the cost to the client.

BVP/PIPS is now being used to successfully deliver commodities, professional services, non-construction services, and Information Technology (IT) services.

Kashiwagi (2011) argues that BVP/PIPS is a process/structure to deliver services. It changes the procurement agent's role from being the guardian over the award of a contract, to a facilitator of the delivery of services. The new role of facilitator starts when a user has a requirement, and ends when the service has been delivered. Instead of being a procurement process, it assists in the development of an intent of the client by expert vendors, identifies the best value vendor (most value for the lowest price), assists the best value vendor to determine if they can meet the intent of the client, and then ensuring that the vendor can deliver on their proposal (Kashiwagi, 2011). The BVP/PIPS has three phases: selection, pre-award, and management of the project risk. The selection phase has five filters: past performance information, competitive ability to manage and minimize project risk, interview of key personnel, prioritizing the vendors, doing a dominance check to ensure that the best value vendor is the best value.

3

Kashiwagi (2011) argues that BVP/PIPS differs from other procurement and risk management systems because it minimizes subjective decision making of the client's experts. It forces the vendors to compete based on value (quality risk management capability and price.) By making the assumption that the vendor is an expert, and disciplining the client's representatives to follow this structure, the client's representatives do not make any technical decisions or judgments on the vendors. If a vendor is dominantly better (easy to see, get a consensus dominant rating, or a non-technical reason why they are dominantly better), they have provided information that clearly shows their dominant performance. If not, the process will be followed, and the best value for the lowest price vendor shall be identified.

In BVP/PIPS, the client's representatives assume the vendors are experts through the selection process, then assume the best value vendor is not an expert in the pre-award phase to minimize the risk of the vendor. The paradigm is to minimize the need for technical decision making in the selection process, and maximizing the need for the best value vendor to prove they are an expert in the pre-award phase (Kashiwagi, 2011).

The philosophy of BVP/PIPS assumes that the suppliers are the experts and the client is the non-expert. This notion implicates that the client (in this case a PCA) cannot understand or grasp ex-ante all possible ways to solve its needs. The client is not fully aware of all its preferences, due to the fact that is has incomplete knowledge on what is possible in the market (and what is not possible). In the philosophy of Best Value Procurement the suppliers each set their own performance level, without the client indicating which performance level it wants (the client does not know what is possible).

Having a client set all criteria, standards, norms and specifications leads to getting more or less the same bids from the different vendors. The advantage of this is that all bids can be easily compared. The downside however is that there is (almost) no differentiation between the bids. Another major disadvantage is that, because of the fact that the vendor is the expert en the client the non-expert, some possible solutions that the client could have gotten are left out up-front (because the client did not see it).

BVP/PIPS seems to ignore most of the rules and demands on evaluation criteria based on the literature; the major exception being the alignment of evaluation criteria to the project goals. In the philosophy of Best Value Procurement the performance level is set by the contractor and the client is not aware of its preference. It is therefore impossible to evaluate a tender without having knowledge of all the other tenders. These scoring functions are known as relative scoring functions. The problem of relative scoring functions is that the overall ranking of tenderers can be dependent of another tenderer and that a phenomenon called rank-reversal can take place (Chen, 2008).

Application of BVP/PIPS in 6 Dutch Case Studies
Rijkswaterstaat is the executive arm of the Dutch Ministry Infrastructure and Environment. On behalf of the Minister and State Secretary, Rijkswaterstaat is responsible for the design, construction, management and maintenance of the main infrastructure facilities in the Netherlands. To help solve the countries' congestion problems, the Dutch Ministry of Infrastructure and Environment has identified 30 major bottlenecks, which need to be (partly) resolved by May 1, 2011. The main reason for using BVP/PIPS is that the procurement of Design and Build-contracts usually leads to high transaction costs (efforts of all possible suppliers) and long tender procedures. In 2009 the tender capacity in the Dutch market was limited. Therefore suppliers have asked Rijkswaterstaat to develop a procurement strategy heavily based on quality to lower the transaction costs and shorten the tender procedure (Van de Rijt & Witteveen, 2011). As a government agency Rijkswaterstaat has to follow the

4

European legislation on public works. The tender process needs to be transparent, non-discriminatory and objective.

Rijkswaterstaat has adopted BVP/PIPS for 16 of the 30 bottleneck projects in order to avoid transaction costs. For each of the projects very concrete project goals were set. While designing the process, the intention was to stay as close to the original PIPS methodology (as developed by Dean Kashiwagi) as possible, with a few adaptations. Van de Rijt & Witteveen (2011) have summoned up the 11 differences with the "pure" methodology. The philosophy when applying the adapted methodology however was still completely intact: it was aimed at finding the highest quality vendor within the budget (like in the original methodology as developed by Dean Kashiwagi), using the risk assessment plan, valued added plan, planning, interviews and price.

- The qualitative criterion Risk Assessment Plan was meant to find out in which way the vendors minimized the risks (mostly risk they do not control) in order to realize the project goal.
- The qualitative criterion Value Add Plan was meant to find out in which way the vendors could add value in order to realize the project goal.
- The qualitative criterion Schedule was meant to find out in which way the vendors could identify the plan from beginning to end in relation to the risks and value adds
- The qualitative criterion Interviews was meant to find out in which way the key people of the vendors could understand their plan from beginning to end.

As one sees, the project goals are central to the evaluation. These qualitative criteria are kept at this rather abstract level in order to provide the vendors as much freedom as needed. This will lead to maximum differentiation between the vendors and the possibility to be pleasantly surprised by a vendor's bid.

For each of these qualitative criteria a vendor could score from a "2" to a "10". There were no ex-ante guidelines of what would lead to a "2" and what would lead to a "10". This would assume that the client could indicate what it expected (where in BVP/PIPS the assumption is that this cannot be identified ex-ante).

To safeguard to objectivity, Kashiwagi's concept of "dominant information" was used to evaluate the different bids. This means information that is simple and easy to observe: different ratings could only be given if there was dominant information that there was a qualitative difference in the bid. Like in the BVP/PIPS process, each team member rated the Risk Assessment plans, the Value Added plans, the schedules and the interviews individually and independently, after which all individual scores were discussed in the team. This team needed to come to a consensus score. The extra "safeguard" (compared with the "pure" methodology) was that for the Risk Assessment plans, Value Added plans and scheduling two teams were installed. This way the process consisted of the following 3 steps:

- Each team member rated the vendors individually on all qualitative criteria (risk assessment plan, value added plan and schedule); with the project goals as a reference
- Coming to a consensus score in a team:
    - The 5 team members of team A came to a consensus score for each of the vendors on each of the qualitative criteria (again with the project goals as a reference)
    - The 5 team members of team B came to a consensus score for each of the vendors (parallel to team A) on each of the qualitative criteria
- Using the consensus scores of team A and of team B a "final" score for each of the vendors (for each criterion) was reached

5

By having this 3-step-approach using dominant information, objectivity can be realized. This is a different way of realizing objectivity then by having very detailed and prescriptive criteria.

Unlike in Kashiwagi's system relative scoring was not used. The bids were assessed by comparing them in quality, but the actual score of a specific bid was done in an absolute way. In the Dutch infrastructure sector bigger public clients have adapted a specific way to combine price and quality into best value (PSI Bouw, 2007), where all "quality" criteria are "transformed" into "fictitious" Euros. To calculate which vendor has the most economically advantageous tender, the amount of "fictitious" Euros scored on quality is deducted from the vendor's budget. E.g. : for a € 100 mln project, the maximum (fictitious) deduction is € 70 mln (=70%). This would lead to a fictitious price for this vendor of € 30 mln. For each criterion, a vendor could get a deduction on its price (when the grade on the quality criterion is more than a "6") or there could be an addition to the price (when the grade on the quality criterion is lower than a "6"). See table 3.

E.g.: if RAVA plans counts for 20% in the ranking of a € 100 mln project, the maximum deduction (resulting from interviews) would be € 20 mln. A score of "7" on interviews would lead to a deduction of € 5 mln.

Grade	% of maximum value
10	100
9	75
8	50
7	25
6	0
5	-25
4	-50
3	-75
2	-100

Table 3: scoring system

Results of the tender
An evaluation of the tender process was commissioned by RWS (Andersson Elffers Felix, 2010). The study consisted of an analysis of existing documentation, and interviews with the market players involved as tenderers for the packages, as well as with those involved within RWS. In addition, conversations were held with market players who have not participated in the tendering process, and with industry organisations. Market players support this method of award, scoring it with an 8 (out of 10). During the interviews most market players actually advocated an even greater weight being placed on quality in the overall assessment. Only a limited number of interviewees are in favour of putting a slightly greater weight on price - 40% or 50% instead of 30%.

Rijkswaterstaat is also positive about the approach. Greater weight has indeed been put on quality, but not as much as intended. That may be due to the tenders lacking points of distinction or to the assessment by Rijkswaterstaat. Interviewees indicated that extreme scores are hardly ever given; there is a natural tendency to refrain from this. Incidentally, this proves better than expected when the variation in assessment is compared for each part of the package. Table 4 shows the minimum and maximum scores for each of the evaluation criteria for the six projects (A-F).

6

	A		B		C		D		E		F	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
Risk assessment	7	3	8	3	7	5	8	5	7	4	7	3
Value added	8	5	9	4	8	6	8	5	7	4	9	6
Schedule	7	5	7	5	8	6	8	6	8	4	8	7
Interviews	9	4	9	4	8	5	8	5	9	3	9	6

Table 4: minimum and maximum scores in the Fast Track Projects

Conclusion and suggestions for further research

The results of the tenders show that it is very well possible to use qualitative scales without further desegregation. The project goals however should be very clear, since they form a central role in Best Value Procurement/PIPS. The results also show that BVP/PIPS enables market parties to differentiate themselves from their competitors and ensures that the PCA's get value for money, As long as the evaluation process is thorough the objectivity of the process can be guaranteed. Objectivity, non-discriminatory and transparency can be realized, even if the client cannot exactly state his preferences.

References

Andersson Elffers Felix (2010), Innovatieve marktbenadering
http://www.rijkswaterstaat.nl/over_ons/zakendoen_met_rws/index.aspx
Belton, V., Stewart, T.J. (2001), Multiple Criteria Decision Analysis: an Integreted Approach. Springer.
Chen (2008) An economic approach to public procurement; Journal of Public Procurement, volume 8, issue 3, 407-430
Dreschler, M. (2009), Fair competition, How to apply the "Economically Most Advantageous Tender (EMAT) award mechanism in the Dutch construction industry
Kashiwagi, D.T. (2009), Best Value PIPS/PIRMS, PBSRG
Kashiwagi, D (2011) Case Study: Best Value Procurement/Performance Information Procurement System Development in Journal for the Advancement of Performance Information and Value, vol 3, nr1, p 12-45
Leeuwen, M. van (2011); Using Best Value PiPS Procurement in Europe, Need for Compromise? Journal for the Advancement of Performance Information and Value, vol 3, nr1, p 56-71
Mateus, R., Ferriera, J.A., Carreira, J. (2010), Full disclosure of tender evaluation models: Background and application in Potuguese public procurement, Journal of Purchasing & Supply Management, 16, 206-215
Palaneeswaran, E. and Kumaraswamy, M.M. (2000), Contractor selection for design/build projects, Journal of Construction Engineering and Management, 17, 315-327
PSI Bouw (2007); Gunnen op Waarde, hoe doe je dat? Rapport in het kader van het PSIBouw-programma Gunnen op waarde (in Dutch).
Rijt, J. van de & W. Witteveen, C. Vis & S. Santema (2011); Best Value at the Directorate-General for Public Works and Water Management in The Netherlands; in Journal for the Advancement of Performance Information and Value (vol 3, nr 1.), p 90-101