The American Society of
Mechanical Engineers

Reprinted From
**PD-Vol. 79, Proceeding of The Engineering Systems
Design and Analysis Conference
Editors: B. D. Czejdo, B. Trousse, I. I. Esta, and B. Shirazi
Book No. G00997 - 1996**

An Expert System for Construction Tendering Process

Waldemar W. Koczkodaj*
Computer Science, Laurentian University, Sudbury, Ontario, Canada P3E 2C6
icci@nickel.laurentian.ca (or waldemar@ramsey.cs.laurentian.ca)

Wojciech Trochymiak†
Institute of Road and Bridges, Faculty of Civil Engineering
Warsaw University of Technology, Al. Armii Ludowej 16, 00-637 Warsaw, Poland

Sudbury in Canada and Warsaw in Poland, April 1, 1996

Abstract

Keywords: project proposal performance evaluation, tendering, experts' opinions, pairwise comparisons, judgement inconsistency

This paper presents a flexible expert system for the evaluation of tenders based on overall aspects of performance. In particular, both tangible factors (such as technical parameters) and intangible factors (such as evaluation of informal relations or environmental assessment) are included in this approach. Considering the complexity of the problem, a model of hierarchical structure is expected. The pairwise comparisons method (introduced in [15]) glues together performance measurements which may take place at many levels. The consistency-driven approach (introduced in [8]) allows one to define conceptual models of tendering processes which are flexible (no fixed list of criteria is assumed) and adaptable to local environments and conform to the local building code requirements.

1 Construction tendering basics

Bridges, roads, buildings, and other civil engineering constructions are often financed by public funds. Selecting a tender (a construction company or a project design proposal) takes place, in most cases, by a public bidding. It is a complicated process which is mostly based on intuition since there is no theoretical base or consistent method of predicting the best bid. It is not uncommon for the evaluation panel to arrive in a deadlock situation when a part of the panel favours one solution because of certain criteria while the other part insists on another solution since, according to their opinion, it scores better on different criteria. The decision making process nearly always involves some kind of constituency in modern democratic societies. We have various boards of governors or directors, committees, task groups, city councils, panels of experts, and individuals, each with an agenda. Heated discussion and various ways of dispute, reasoning, and argumentation take place to arrive at certain decisions. Most constituencies have worked out precise and practical policies for running meetings in an orderly and effective way. What we lack, however, is a device for drawing solid consistent conclusions and all too often the loudest individual wins! Unfortunately loudness does not necessarily go along with wisdom. Casual thinking does not work well in predicting complex outcomes. Casual thinking is partial, fragmentary, and is not an effective way to measure intangibles. In the decision making process many factors must be considered simultaneously and with about the same degree of importance, therefore an approach with more finesse is necessary to obtain a clear and unambiguous conclusion. It has been shown by numerous examples (see Section 4 for more detailed discussion) that the pairwise comparison method can be used to draw final conclusions in a comparatively easy and elegant way. The brilliance of the pairwise comparison could be reduced to a common sense rule: consider two factors at a time if you are unable to handle more than that.

The main goal of tendering, which is usually organized by the construction investor, is the selection of the most suitable tender from the public point of view. Through a public bidding we try to achieve:

- the setting of common input constraints for potential suppliers or/and constructors,
- the selection of the best tender based on tangible and intangible but constant (during the entire bidding process) criteria which allow us to compare the proposed offers,
- a minimization of the influence of informal interests on

---

[*] partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)

[†] partially supported by the General Directorate of Public Roads

selection of an offer thanks to the application of a strict selection process.

One of the most instrumental conditions of a fair public bidding is the necessity of a precise scoring system of all criteria and their preferences in the tendering documents to be used during the selection process by the selection panel.

2 Basic concepts of the pairwise comparison method

The pairwise comparison methodology introduced by Thurstone in 1927 (see [15]) can be employed as a powerful theoretical framework for the evaluation of civil engineering tenders. Some of the notable past applications of national importance are the evaluation of a proposal to build nuclear power plants in Holland (rejected by Dutch Parliamentary Committee on the basis of the pairwise comparison method and the evaluation of transportation system proposals in Sudan (for details and more examples of applications see [11, 13, 3, 12, 1, 5]).

The practical and theoretical virtue of the pairwise comparison methodology is its simplicity. The goal of pairwise comparisons is to establish the relative preference of two criteria in situations in which it is impractical (or sometimes meaningless) to provide the absolute estimations of the criteria. To this end, an expert (or a team of experts) provides relative comparison coefficients $a_{ij} > 0$, which are meant to be a substitute for the quotients s_i/s_j of the unknown (or even undefined) absolute values of criteria $s_i, s_j > 0$. The quotients s_i/s_j are also sometimes called *relative weights* in the literature.

$$\mathbf{A} = \begin{vmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{vmatrix}$$

where a_{ij} expresses an expert's relative preference of criteria s_i, over s_j.

Coefficients a_{ij} are expected to satisfy some natural restrictions (e.g. $a_{ii} = 1$, $a_{ij} \cdot a_{ji} = 1$). For the sake of our exposition we define the pairwise comparison $n \times n$ matrices simply as square matrices $A = (a_{ij})$ such that $a_{ij} > 0$ for every $i, j = 1, \ldots, n$.

A pairwise comparison matrix A is called *reciprocal* if $a_{ij} = \frac{1}{a_{ji}}$ for every $i, j = 1, \ldots, n$ (then automatically $a_{ii} = 1$ for every $I = 1, \ldots, n$). Even a stronger condition seems natural. A pairwise comparison matrix A is called consistent if $a_{ik} = a_{ij} \cdot a_{jk}$ for every $i, j, k = 1, \ldots, n$. While every consistent matrix is reciprocal, the inverse in general fails. Consistent matrices correspond to the ideal situation in which there are exact values $s_1, \ldots, s_n$ for criteria. The quotients $a_{ij} = s_i/s_j$ form a consistent matrix. Conversely, the starting point of the pairwise comparison inference theory is Saaty's theorem (see [14]) which states that for every $n \times n$ consistent matrix $A = (a_{ij})$ there exist positive real numbers $s_1, \ldots s_n$ such that $a_{ij} = s_i/s_j$ for every $i, j = 1, \ldots, n$. Such vector $s = (s_1, \ldots s_n)$ is unique up to a multiplicative constant.

Code	Definition of intensity or importance	Application
1	Equal or unknown importance	Two criteria equally contribute to the objective or lack of knowledge to compare them
2	Weak importance of one over another	Experience and judgments slightly favour one criterion over another
3	Essential or strong importance	Experience and judgments favour one criterion over another
4	Demonstrated importance	The criterion is strongly favoured and its dominance is demonstrated in practice
5	Absolute importance	The highest affirmation degree of favouring one criterion over another
2.4, etc.	Intermediate judgments	When compromise is needed

Table 1: Comparison scale

The challenge posed to the pairwise comparison method comes from the lack of consistency of the pairwise comparison matrices which arise in practice (while as a rule, all the pairwise comparison matrices are reciprocal). Given an $n \times n$ matrix A which is not consistent, the theory attempts to provide a consistent $n \times n$ matrix B which differs from matrix A "as little as possible". One of the possible solutions to this problem was proposed by Saaty (see [14]). An alternative solution is presented in [4] as geometric means. There is no visible superiority of one method above the other (contrary what some papers claims) as shown in [7] by a Monte Carlo experiment on 10,000,000 cases. Let $s = (s_1, \ldots, s_n)$ be the eigenvector of A corresponding to σ, the largest eigenvalue of A. Furthermore, vector s can be chosen to have all coordinates positive.

How can we establish fair weights? Is there any theory to help us? The weighting classification needs to be done on fair basis for every criteria which ought to have its share in contributing to the overall judgment. A fair solution is to compare all criteria in pairs using, for example, a scale from

1 to 5 presented in Table 1.

3 An example of a construction tendering model

A practical model of a construction tendering process needs to be as flexible as possible (see also conclusions). Presenting any model here is risky since a less careful reader may conclude that this model is not suitable for him/her. One may always discount any model as irrelevant, however, leaving a reader without any practical application of the presented framework would be considered as cruel and unusual punishment therefore a compromised solution is proposed.

The model presented in Fig. 1 is applicable to both unrestricted and restricted bids as well as to two-stage biding. Table 2 contains a set of selected criteria most frequently used in construction tenders. It includes (amongst others) criteria discussed in [10]. This list is not exhaustive and does not pretend to be complete. It is worthwhile to note that the authority issuing a tender can select arbitrary criteria for each type of bid. They can also scale (or weight) particular criteria, depending on the kind (or extent) of works, the required potential of the contractor, or the necessary level of technology.

The criteria have been divided into five main groups:

- the price for which the bidder is prepared to carry out works that are the subject of the tender,
- proposed design - this criterion is not present in cases when the authority issuing the tender already has its own design (e.g. of a building, road, bridge). Presence of this criterion allows the selection panel to assign preferences to such criteria as aesthetics, ecological awareness, use of innovative technology, durability assessment, and to assess how easy it will to maintain and service a construction,
- assessment of the reliability of the bidder - the proposed set 2 of criteria has been divided in this case into two groups enabling the evaluation of the bidder to be made from the point of view of personnel qualifications, experience gained carrying out other contracts, necessary plant and laboratory equipment and - on the other hand - to enable the assessment of the bidder to reflect whether he is able to carry out the task from the financial point of view (e.g. having necessary credit for initial purchases of materials and to pay the personnel),
- preference applied - legislation in force in the area of public procurement allows in many countries various preferences to be made, allowing to prefer definite kinds of contractors (e.g. domestic firms, or firms from the neighbouring countries, or well proven, reliable contractors, who apply domestic materials and technologies),
- proposed construction period.

Tender evaluation criteria

- Project
    - Price quotation (*price*)
    - Proposed design (*design*)
        - Quality of the proposed concept (*quality*)
        - Durability of the prop. concept (*durab*)
        - Undefined factors (*u-fact*)
            - Environmental factors (*envir*)
            - Technological innovation (*new-tech*)
        - Reliability of the proposed concept (*reliab*)
    - Assessment of contr. reliability (*reliab*)
        - Technical reliability (*relia-t*)
            - Experience (*exp-ce*)
            - Personnel qualifications (*staff*)
            - Own plant (*plant*)
            - Laboratory equipment (*lab*)
            - Quality control system (*q-syst*)
        - Economic reliability (*relia-e*)
            - Profitability (*profit*)
            - Financial liquidity (*liquid*)
            - Debt (*debt*)
            - Credit ability (*credit*)
            - Participation in other contracts (*o-contr*)
    - Applied preferences (*pref*)
        - Preference for domestic firms (*dome-f*)
        - Territorial preferences (*territ*)
        - Experience in cooperation (*cooper*)
        - Informal relations (*relation*)
    - Proposed construction period

Table 2: Criteria often taken into consideration in the evaluation process

The model in Fig. 1 assumes that in the submitted tender the bidder is requested to present his own project design. The design is assessed together with other factors such as the price, construction period, producing documents supporting the reliability of the bidder and preferences applied. In the general case the project design does not need to be submitted. The criteria in the presented model are fairly general, but they embrace the entire range of problems occurring in construction tenders.

It is hard to discuss all comparisons therefore only the group of DESIGN is shown by Fig. 2. This criteria group assesses alternative concepts of constructions (e.g. a bridge, road, or building). Figure 2 demonstrates a matrix with relative comparisons. A scale of 1 to 5 (and its inverse 1/5 to 1) is used. In the represented case the highest importance has been assigned to *quality* (quality of the proposed concept)

	quality	durab	u-fact	reliab
quality	**1.00**	1.30	2.00	4.00
durab		**1.00**	1.50	3.00
u-fact			**1.00**	2.00
reliab				**1.00**

Figure 2: Relative judgments for the DESIGN group of criteria

because the construction quality is usually of a great importance for both an investor and user. It has been assessed to 1.3 in comparison to the durability of the proposed concept (*durability*) of the construction which is a second criterium as far as importance is concerned. Construction durability is factor related to the expected period of time which the construction should last.

It is assumed that *Undefined factors* are less important for the investor. They allow us to assess such evaluation criteria as aesthetics, disturbance of the environment, and technological innovations. The assessment of importance of quality factors against durability is set to 2 which is according to Table 1 *Weak importance of one criterion over the another*. *Durability* against *undefined factors* (which are in the model in Figure 1 subdivided into two categories) is assessed to 1.5 (a compromised evaluation between 1 and 2). The last criterium in the DESIGN group is *reliability of the proposed concept (reliability)*. It is related to the safety of the proposed construction. Its importance is lower since all of the proposed designs are expected to meet a certain specified minimum (otherwise they would be disqualified from the bidding). *Quality* is far more important than *reliability* (this is the meaning of code 4 in Table 1). *Durability* against *reliability* is set to 3.0 and *undefined factor* to 2.0 when compared to *reliability*. Finally the *undefined factor* have been assessed as more important than *reliability* which is reflected by 2.00 in the last column of row three.

Bold face 1's on the main diagonal are arbitrary values since they represent a relative ratio of a criterium against itself. Values below the main diagonal do not need to be entered by the user. They are reciprocal to the corresponding values in the upper triangle (for details see matrix **A** in Section 3).

4 Inconsistency analysis

A careful reader may be curious (if not suspicious) about how one could arrive with values such as 1.30 or 1.50 as relative ratio judgements. In fact the initial values were different but they have been refined and the final weights have been calculated by the consistency analysis. It is fair to say that making comparative judgments of intangible criteria (e.g. informal relationship with the contractor) involves not only imprecise or inexact knowledge but also inconsistency in our own judgments. The improvement of knowledge elicitation by controlling the inconsistency of experts' judgments (also called *consistency-driven approach*) is not only desirable but absolutely necessary.

In practice, inconsistent judgements are unavoidable when at least three factors are independently compared against each other. For example, let us look closely at the original ratios of the first three criteria in Figure 2: *quality* (for short A), *durab* (denoted by B), and *u-fact* (referred as C). The original assessment of A against B was 2, B against C was assessed as 2. One may sense a problem since the ration of A to C was also 2. It is a quite frequent case which we unintentionally tend to experience: everything seems to be important!

The above reasoning may illustrate the need for inconsistency analysis. From $\frac{A}{B} = 2$ and $\frac{B}{C} = 2$ we can infer that $\frac{A}{C} = 2 \cdot 2 = 4$. However, assuming that $\frac{C}{A} = 2$ and $\frac{B}{C} = 2$ we infer that $\frac{A}{B} = 1$ which is different from our original assessments. In fact we do not know which assessment was incorrect. In particular (a frequent case in practice) each original assessment might have been (an usually is) *just a little inaccurate*.

The consistency factor (cf) is the minimum of $|1 - \frac{2}{2 \cdot 2}|$ and $|1 - \frac{2-(2 \cdot 2)}{4}|$ which is 0.50 and rather high. Since we are not in a position of saying which ratio is incorrect, all three judgements must be reconsidered before any further calculations (e.g. of the final weights) can take place. For details related to consistency analysis see [8, 6]. In our situation a compromised solution have been employed: ratio *durab* to *u-fact* has been decreased to 1.5 causing the decrease of the highest inconsistency to drop to 0.33 for anther combination of three criteria: *quality*, *reliab*, and *durab*! In fact the inconsistency exists if more than two judgments are involved (see [8] for details). Having already experience with the former case we decrease 2 (the ration of *quality* and *durab*) to a compromised value of 1.5 receiving a very small inconsistency factor 0.11 which is below a acceptable threshold 0.33 (see [8] for details). There is no practical reason to continue decreasing the inconsistency (only it high value is harmful; a very small value may indicate racing the data rather than entering the honest assessments). However, for the demonstration purposes we decreased the last value of 1.5 to 1.3 receiving near-zero inconsistency: 0.03 which should be considered as rather an unusual (and suspicious) case.

An interesting case takes place when we start with a different set of the initial judgements in the dashed boxes of Figure 2. Let us assume (for the sake of discussion) that they are: 1.00, 2.00, and 2.00. The dashed triad of judgements is obviously consistent since indeed 1.002.00 = 2.00 (see Section 2; by the definition the value of the corner element should be equal to the product of the remaining two elements). This means that the value of *cf* (the consistency factor) is 0.00 but only for the dashed triad. The consistency factor of the triad marked with regular line boxes (that is with values 1.00, 4.00, and 3.00) has inconsistency 0.75. Let us try to change the value of 1 in this triad (according to Table 1, equal or unknown importance) to a compromised value of 1.5. Now the highest inconsistency moved to the dashed triad where we may, in turn, chose to decrease value of 2 in the lower triangle corner to 1.5 (an attempt of trying the same alteration with the other 2 results in increasing the inconsistency). As a final touch we may change 1.5 to 1.3 the same way as it has been described earlier in this Section.

The inconsistency analysis may look complicated but the software developed for this analysis allowed a more practical approach. By decreasing or increasing a certain value (the most inconsistent combinations are provided by the software) one develops a very good orientation quite quickly.

All the above computations including the final weights are done by a software. It is not important to address all mathematical aspects of getting the final weights but the eigenvector method (see [14, 8] for details) can be used to obtain results illustrated in Figure 3.

There is one interesting observation from looking at the running total. The first ten criteria constitute over 85% of the score. In other words if tender X scores say 55% on the first 10 criteria and the leader's score (on these 10 criteria) is say 72% then there is no need to evaluate the remaining 11 criteria for X. Simply, X is unable to win with the leader since the difference between them is 17% (that is 72%-55%). The practical approach would be to evaluate all tenders on the first 10 criteria (it may be a substantial time gain) and take only those who are trailing the leader with less then 14.33% (that is 100%-85.67%).

Using a more precise consistency-driven approach may look complicated at the first glance. It may be particularly visible when it comes to making the comparative judgments. A statistical experiment (see [9]) proves that the precision gain is substantial (as much as 300% when it was tested on 100 students for estimation of lengths of randomly generated bars). (Results and software, *The Convincer*, are available upon request by email).

5 Conclusions

The bidding process may be modeled in many different ways. A good model should, however, contain both tangible and intangible factors. Identification of major criteria is one of essential components of building such a model. They may strongly depend on the specifics of the case (e.g. some types of bridges may be excluded) such as: the environment, the local building code, etc. Our approach allows the investor to create its own list of criteria without any limitations.

Probably one of the most important alterations to the model is the price. As a decisive factor (but also the most political) it may be removed from the model for the decision by the evaluation panel. In other words, the model would provide the evaluation model with the assessments of tenders as a vector of weights. These weights will be applied to evaluation of each individual proposal. The final product of the evaluation will be a list of tenders with the overall score (received by a vector product of weights and corresponding independently made evaluations for each criteria) plus the price.

The inconsistency analysis and the refinement of judgements is supported by software. The interpretation of the final results and public acceptance of them may require substantial amounts of social and political effort and time, but it is worth it. The precise weights and a structured evaluation procedure contribute to the selection of the best tender in a comparatively shorter period of time with less risk for any grievance. Sometimes it may be wise to engage a computer consultant (usually for fraction of a percent of the construction cost) when a project is of provincial, national, or political importance for running the software and guidance in the consistency-driven approach.

The presented results have been obtained by *The Concluder* system and are available (together with the presented model) to interested readers from the authors upon email request. It has been released to the public domain for everyone's benefit. *The Concluder* runs under MS Windows^{TM} on personal computers and does not require any specific mathematical knowledge.

References

[1] Bolger P.M., Duszak Z., Koczkodaj W.W., Mackasey W.O. 1993, Ontario Abandoned Mine Hazards Prioritizing - an Expert System Approach. In: Proceedings of the 15th Annual Abandoned Mine Land Conference, Jackson, Wyoming, September 13-15, pp. 370-388, 1993.

[2] Caffi, C., Janney, R. W. *Introduction: Planning a bridge*, Journal of Pragmatics, Vol. 22, No 3/4, pp. 245-260, 1994.

[3] Ching-Lai, H., Kwangsun, Y., Multiply Attribute Decision Making. Methods and Applications, Springer-Vorlage, Berlin Heidelberg, 1981.

[4] Crawford, G., *The Geometric Mean Procedure for Estimating the Scale of Judgement Matrix*, Mathematical Modelling, Vol. 9, pp. 327-334, 1987.

[5] Duszak Z., Koczkodaj W.W., Mackasey W.O., *Towards Better Abandoned Mine Hazard Prioritizing - An Expert*

System Approach, in The Challenge of Integrating Diverse Perspectives in Reclamation. Proceedings of the 10th National Meeting of ASSMR, Spokane, Washington, pp. 577-589, 1993.

[6] Duszak, Z., Koczkodaj, W.W., *The Generalization of a New Definition of Consistency for Pairwise Comparisons*, Information Processing Letters, Vol. 52, pp. 273-276, 1994.

[7] Herman, M., Koczkodaj, W.W., *Monte Carlo Study of Pairwise Comparisons*, Information Processing Letters, 57(1), pp. 25-29, 1996.

[8] Koczkodaj, W.W. *A New Definition of Consistency of Pairwise Comparisons*. Mathematical and Computer Modelling, 18(7), 79-84, 1993.

[9] Koczkodaj, W.W., *Statistically Accurate Evidence of Improved Error Rate by Pairwise Comparisons*. Perceptual and Motor Skills, 82, 43-48, 1996.

[10] Majewski, Bohdan, *One of the longest wood bridges in Poland over the Bug river (392m long) will be replaced by a more durable steel bridge*, Bulletin of the Public Roads (Biuletyn Informacyjny Drogownictwa), General Directorate of Public Roads. 1/94, p. 18-20, 1994 (in Polish).

[11] Nijkamp, P., Rietvield, P., Voogd, H., *Multicriteria Evaluation in Physical Planning*, North-Holland, Amsterdam, 1990.

[12] Saaty, T.L., *The Sudan Transport Study*. Interfaces, Vol. 8, No. 1, pp. 37-57, 1977.

[13] Voogd, H., Multicriteria Evaluation for Urban and regional Planning. Pion Ltd., London, 1983.

[14] Saaty, T.L. *A Scaling Method for Priorities in Hierarchical Structure*. Journal of Mathematical Psychology, Vol. 15, 234-281, 1977.

[15] Thurstone, L.L., *A Law of Comparative Judgments*, Psychological Reviews, Vol. 34, 273-286, 1927.

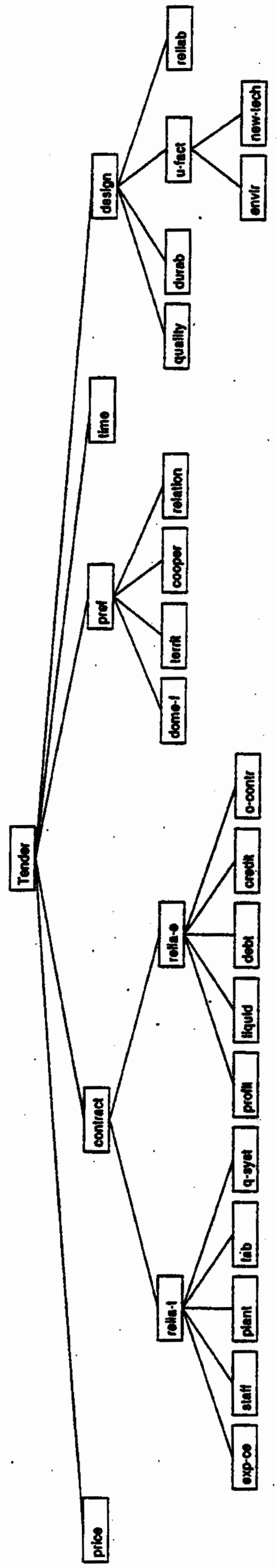
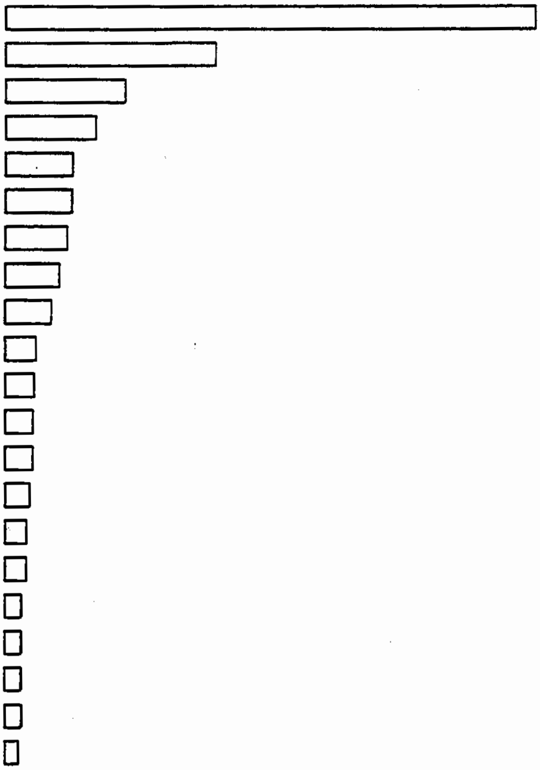


Figure 1: An example model of a construction tendering

	Criterium name	Weight	R-tot	Bar graph representation
1.	Price quotation	36.10%	36.10%	
2.	Proposed construction period	14.16%	50.26%	
3.	Quality of the project. concept	7.90%	58.16%	
4.	Durability of the prop. concept	6.00%	64.16%	
5.	Preference for domestic firms	4.44%	68.59%	
6.	Experience	4.43%	73.02%	
7.	Personnel qualifications	4.12%	77.14%	
8.	Experience in cooperation	3.56%	80.70%	
9.	Environmental factors	2.98%	83.68%	
10.	Reliability of the prop. concept	1.99%	85.67%	
11.	Credit ability	1.86%	87.53%	
12.	Participation in other contracts	1.78%	89.31%	
13.	Debt	1.78%	91.09%	
14.	Financial liquidity	1.55%	92.64%	
15.	Own plant	1.33%	93.97%	
16.	Laboratory equipment	1.28%	95.25%	
17.	Territorial preferences	1.02%	96.27%	
18.	Informal relations	1.00%	97.27%	
19.	Technical innovation	0.99%	98.26%	
20.	Profitability	0.98%	99.24%	
21.	Quality control system	0.79%	100.0%	

Where R-tot means *running total* of the weight

Figure 3: The final weights for the evaluated criteria