Commerce Decisions
A QINETIQ company

White Paper

Effective weighting of criteria as a critical component in delivering the right decision in public procurement

Richard Tottman, Principal Consultant
Commerce Decisions Limited

Executive summary

This White Paper looks at the issues surrounding weighting, what weighting is for, and some methods that can be used to create weights that will increase the likelihood of yielding the best decision.

The process of weighting involves emphasising some criteria more than others. The output of the weighting process is a set of criteria where each criterion has a different impact on the result.

What is it that makes one factor more important to a decision than another? When trying to determine weight, how should the decider go about choosing which criteria are more important? There are many methods available, each with its pros and cons.

Contents

Why weight?

Modern decision-making relies on the identification and balance of a (potentially large) list of criteria. In public procurement, it is necessary to describe the decision-making process to tenderers so that they may best represent themselves to the buying authority and optimise their solution to fit the buyers' needs. The more complicated the purchase is, the more detailed the description of the decision-making process needs to be.

Otherwise, the buyer runs the risk of not evaluating the right content from the tenderers and choosing the bid that is not the one most likely to deliver the right solution at the right cost.

Every decision that is made involves comparison, even if it is against the 'do nothing' alternative. We do this quickly and subconsciously for small decisions, and thoughtfully and deliberately for larger decisions. Choose any subject and you will find yourself weighing up different needs and coming to a judgment.

For instance, when we are deciding which car to buy, is comfort or reliability more important to us? When looking for somewhere to live, would we trade room size for access to facilities? Also, once we have made a judgment about what is more important to us, do we understand why it is important? Some methods will help us decide the weight; other methods will help us decide the weight and understand why the weight was chosen.

When we are making collaborative decisions involving many stakeholders, we need methods to underpin our decision and records to defend the result. If we do not agree the various factors and their relative importance, then chaos, or extensive discussions, can ensue when we are presented with the options and must reach an agreement.

There are two major elements to construct when making a decision. One is the list of criteria; the other is the importance of each criterion to the final decision. This paper deals with how to decide how each of the criteria should affect the outcome.

Why share the 'weight' with the bidders?

So, if we are able to weight the various factors, what is the benefit and, more importantly, what are the risks of openly sharing that information with a tenderer?

The benefit is easy to state: "The increased knowledge of our decision factors will enable the tenderers to focus their response on those factors".

Paradoxically, this is also the reason that authorities often state for not declaring weights: "We will increase the risk that we will be told what we want to hear".

Actually, the declaration that one factor is more important than another gives the tenderer opportunity to focus the effort and optimise the solution on the more important factor. More importantly,it ensures that all the tenderers have the same level of understanding, and the same opportunity to focus on the important factors. The result is that any evaluation of the tenders is likely to offer a more appropriate comparison and the bids will offer capability and costs that are more aligned with the needs of the Authority.

To consider whether the benefit of openly declaring the weights outweighs the risk, the Authority must consider a couple of things.

Firstly, what is the likelihood that it will receive an inappropriate offer from potentially the best supplier? This can happen when very capable and adaptable suppliers can choose the type of solution that they would like to offer. The best supplier can simply choose an option, get it wrong, misrepresent themselves, and not win the contract. Also, a less appropriate supplier can choose an option, guess it right, and win the contract. If the question and its weight is understood, then everyone has the same chance of getting the answer right.

Weighting and the law

The Public Contracts Regulations 2006 state that the contracting authority's invitation shall include the relative weighting of criteria for the award of the contract or, where appropriate, the descending order of importance for such criteria [16. Restricted Procedure (15) (d)].

The Regulations also state that:

"When stating the weightings, a contracting authority may give the weightings a range and specify a minimum and maximum weighting where it considers it appropriate in view of the subject matter of the contract."

In the case of International Ltd v London Borough of Newham1, the court emphasised that the requirement of transparency means that all criteria used by a contracting authority to identify the most economically advantageous tender and their relative importance must be disclosed in advance so that tenderers are aware of them when they prepare their bids.

The ECJ ruled in January 2008 that

"A contracting authority cannot apply weightings and sub-criteria to award criteria set out in tender documentation unless those weightings or sub-criteria have been previously brought to the tenderers' attention".[2]

The omission of detailed weightings from an invitation may therefore invite challenge if the weightings are later used to make a decision. The Court of Appeal in the Newham case ruled that it is not necessary to show that a tenderer would have been successful in winning a contract to be awarded damages, only that the tenderer has experienced a loss, such as bid costs. Therefore, any omission of information from a contracting authority increases the risk that the decision will be challenged.

The continuing development of case law poses a problem for contracting authorities. The regulations stipulate that the Authority may provide order of importance information and ranges of weights rather than the fixed values. It would be reasonable for the tenderers to expect that the actual weights had been fixed before the start of any assessment. What better way to prove that than by publishing them?

Deciding what weight is for

Many groups of stakeholders embark on weighting exercises without carefully considering the consequences of the decisions that they are making. How we design our weights can have a dramatic effect on the outcome.

Imagine that you are able to assess potential candidate solutions using eight criteria. Each criterion has a threshold (minimum acceptable) and an objective (most useful) level. Each criterion is vital to the success of your project, and must be delivered to at least the threshold level.

Tenderers A, B, C and D submit their tenders and are assessed. The raw results are shown in Table 1.

A simple scheme might state that each requirement will be given the same weight, and 50% of the score be awarded for meeting the threshold and another 50% for the objective (call this scheme W1). The results would be as shown in Figure 1.

However, we could state that any solution that does not meet the threshold will be eliminated from the competition. So our weighting scheme can be stated differently. Each question is evenly weighted and 100% of the marks will be awarded for meeting the objective level. 0% will be awarded for scoring the threshold level.

Any solution not meeting the threshold will be eliminated from the competition (call this scheme W2). The results would be different, as shown in Figure 2.

Deeper analysis might show that nine of the requirements are vital to successful operation and analysis shows that any supplier will be able to deliver to our objective level. The remaining requirement (RQT9) is also vital but achieving the objective level for this requirement offers greater usefulness (or benefit) to the buyer

The results (Figures 1, 2 and 3) show very different outcomes for the different schemes. The more the requirements (or needs) are analysed , the more accurately the results from the calculations will reflect the benefits.

A well-designed weighting mechanism therefore serves two purposes: it rewards the solution that best meets the Authority's needs and quantifies the differences between the solutions. A poorly-designed weighting system will do neither.

Raw results

	A	**B**	**C**	**D**
RQT 1	Threshold	Complies with objective	Complies with objective	Non- compliant
RQT 2	Threshold	Threshold	Complies with objective	Complies with objective
RQT 3	Threshold	Threshold	Complies with objective	Complies with objective
RQT 4	Threshold	Threshold	Threshold	Complies with objective
RQT 5	Threshold	Threshold	Complies with objective	Complies with objective
RQT 6	Threshold	Complies with objective	Complies with objective	Complies with objective
RQT 7	Complies with objective	Complies with objective	Threshold	Complies with objective
RQT 8	Complies with objective	Complies with objective	Complies with objective	Complies with objective

Table 1. Raw results in tender assessment

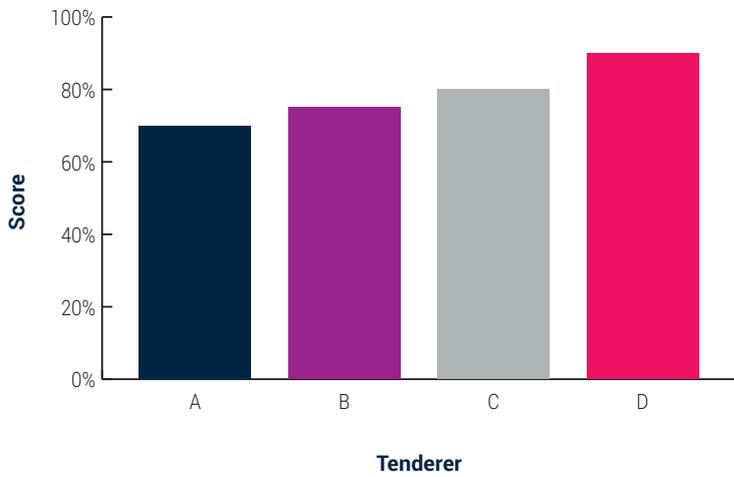**Figure 1. Results using equal weighting (W1 Scheme)**

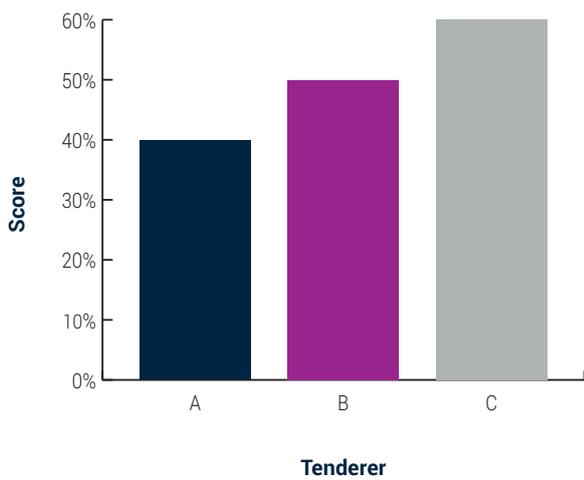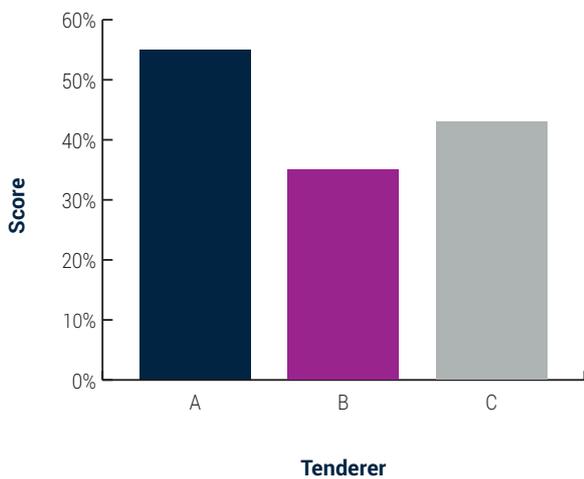**Figure 2. Results using elimination (W2 Scheme)**

Figure 3. Results using deeper analysis (W3 Scheme)

Weighting methods described

There are many ways of producing weights. This paper covers some of those with scientific merit and some of those used in public procurement.

Regardless of the method used, there is a fundamental aspect that can dramatically influence the success of a weighting scheme. When we are trying to distinguish the relative importance of things, it is important that we have an understanding of the things that we are comparing.

For instance, consider an attempt to weight two items: one technical, and the other commercial. Each of us will have an idea about the relative importance of these items, which may be wrong when we find out more information about these criteria. We will all be carrying around baggage about what each of these titles mean. This baggage will be determined by our role and experience.

By elaborating these items we can increase our understanding and make better judgements about the relative importance of them.

Inheritance

Potentially the most dangerous method of weighting, but nevertheless a practice used, is inheritance. This is where the weights (and often the questions) are copied from a previous similar project. The project may be regarded as having a similar requirement and consequently the weights are considered to be similar. This method ignores the needs of stakeholders and does not consider that the risks involved in the new project may be very different from those used previously.

100% subjective method

This method normally relies on the knowledge of the senior stakeholder. The method gives different sections of the evaluation arbitrary percentages based on 'experience'. The different areas of evaluation within those areas are then given percentages and so on until all questions have some percentage assigned. This method is prevalent where the decision makers have a technical project management background.

This method produces a result very quickly. However, the results often do not stand up to scrutiny and it is also difficult to get stakeholder buy-in.

Pair-wise comparison

Pair-wise comparison is a mathematically and academically robust method available for weighting. It involves the comparison of each criterion with all the other criteria. When the number of criteria is large this creates a large number of comparisons.

Total comparisons = n * (n-1)

where n is the number of items to compare. So, for 50 criteria, that is 2450 comparisons! So this method is only really practical for small numbers of criteria.

Normally these comparisons are made using a questionnaire structured in such a way as to make it difficult for the respondent to ensure consistency of comparison between different elements of the set. The results of these comparisons are therefore normally (on the surface) ambiguous. This could be because of 'triangular inconsistency' – e.g. 'A is more important than B'; 'B has the same importance as C'; 'C has the same importance as A' is inconsistent. This result is natural, but designing a questionnaire to allow it happen naturally is very difficult.

Because of this resulting inconsistency, the results of the pair-wise comparisons are processed using advanced mathematical techniques (iterative eigenvalue multiplication) before weights can be calculated. This requires specialist tools and often expensive consultancy.

The fact that pair-wise comparison addresses inconsistency is both a strength and a weakness. The results of the weighting are mathematically robust, but the method of calculation of the weights is opaque and often difficult to understand. This reduces the ability of the stakeholders to agree with the outcome unless they are fully engaged with the process. If a senior stakeholder disagrees with the outcome the only option would be to start again. This is often a time- consuming and costly exercise.

Pair-wise comparison is used extensively in The Analytical Hierarchy/ Network Process developed by Thomas L Saaty.

Scaled comparison

At the other end of the complexity spectrum is a simple 'scaled comparison' method for weighting. It is easily deployed and understood by stakeholders, especially when weighting in a workshop environment. Using this method the stakeholders have the opportunity to understand the process and refine their initial comparisons.

This means the results can be considered less mathematically and academically rigorous. However, because the stakeholders can be more involved in the process and have the opportunity to question and refine the comparisons being made, weights of an acceptable accuracy can be developed with the benefits of enhanced stakeholder involvement and buy-in.

This method is based on the organisation of the criteria into a hierarchy. Having done this, the criteria are divided up into sets of comparison groups of 'peer criteria' (normally siblings of a parent criterion in the hierarchy). A criterion is chosen from each comparison group and the others in the group compared with it. These resulting comparisons are then used to calculate absolute weights in percentage terms (see Figure 4).
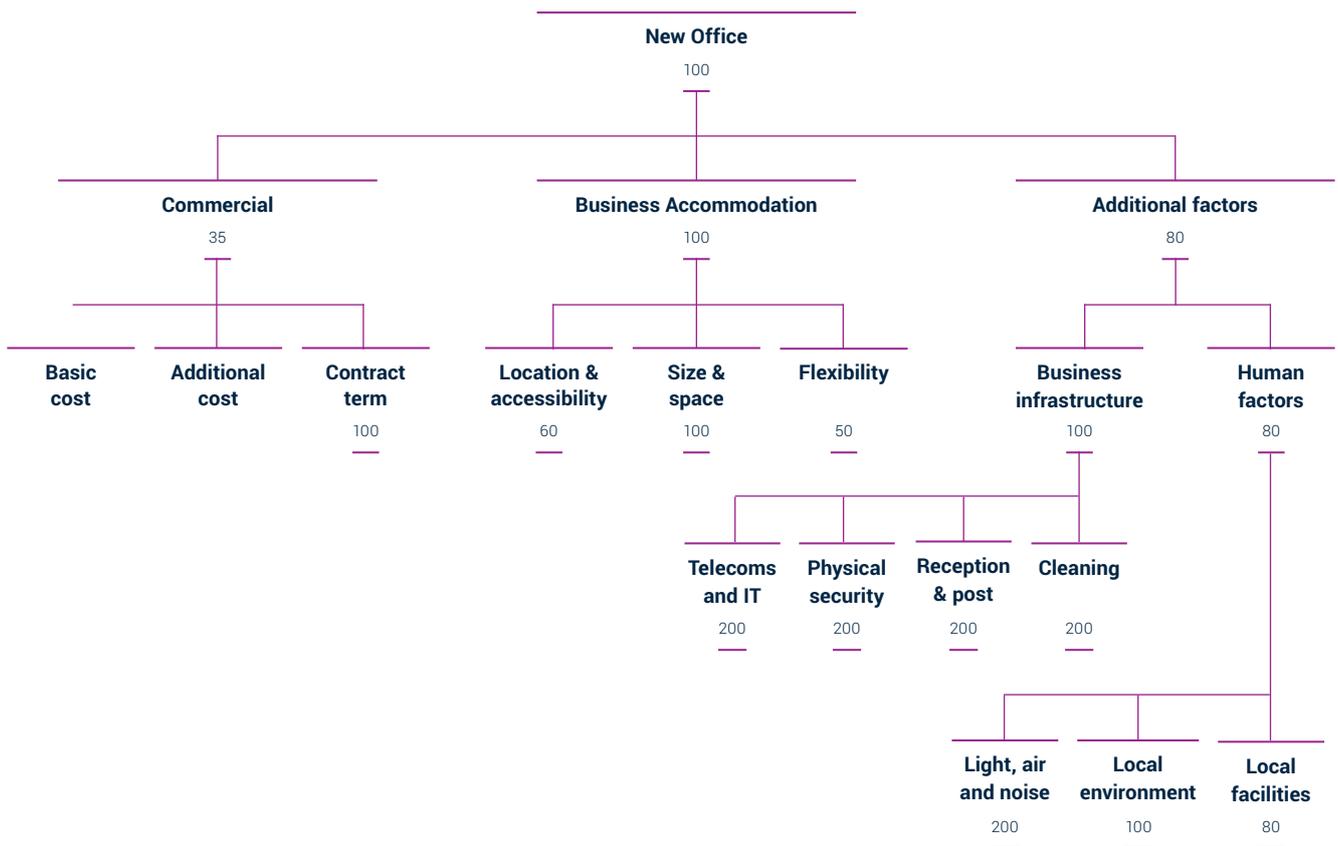
New Office
100

Commercial
35

Business Accommodation
100

Additional factors
80

Basic cost

Additional cost

Contract term
100

Location & accessibility
60

Size & space
100

Flexibility
50

Business infrastructure
100

Human factors
80

Telecoms and IT
200

Physical security
200

Reception & post
200

Cleaning
200

Light, air and noise
200

Local environment
100

Local facilities
80

Figure 4. Weighting comparison groups using scaled comparison

Limitations of the mind

When comparing criteria, it is important to understand the natural limitations on the ability of the human mind to perform the comparisons as a result of language and the mind's ability to differentiate between similar items.

The mind finds it difficult to differentiate between items that differ very little in magnitude and also finds it difficult to judge accurately the relative difference between items that differ by an order of magnitude. For example, the difference in size between an orange and apple can be judged reasonably well, whereas the difference in size between a grape and a watermelon is very difficult to judge accurately.

It has been found that we can make accurate comparisons using natural language alongside numbers, such as 'equal preference', or 'strong preference'. For this reason it can be useful to use normal comparative adjectives when making comparisons. This technique is often used in customer surveys or opinion polls. When doing this, it is necessary to define the allowable comparative adjectives and to set their numerical value. Some scales proposed by the scientific community are given in Table 2.

These scales are derived from experiments using things that can be measured physically, such as brightness of lights or area of shapes. Experiments have shown that different scales yield more accurate results for different physical phenomena. The difficulty with weighting criteria is that there can be no absolute proof as to whether the scale is 'right' or not, because there is no physical comparison. We must therefore look to other ways to gain confidence that the weights are 'right'.

Verbal statement	Balanced[3]	9/9…9/1[4]	AHP scale[5]
Equal preference	1.00	9/9 = 1.00	1
	1.22	9/8 = 1.13	2
Moderate preference	1.50	9/7 = 1.29	3
	1.86	9/6 = 1.5	4
Strong preference	2.33	9/5 = 1.80	5
	3.00	9/4 = 2.25	6
Very strong preference	4.00	9/4 = 2.25	7
	5.67	9/2 = 4.50	8
Extreme preference	9.00	9/1 = 9.00	9

Table 2. Using adjectives for weighting

Objective weighting

It is possible to create a weighting scheme based on objective, real- world values that reflect the impact of a criterion on the project. This is particularly true for procurement projects where, traditionally, time, cost and quality are the key aspects. In this method, each criterion can be compared against a set of stated conditions and an associated weight stated.

An advantage of this method is that the criteria can be defended against each condition. Table 3 shows the start and end of a generic example of such a scheme. For a criterion to have a VERY SIGNIFICANT impact, a failure in the potential solutions would need to have the impact of doubling the cost or time estimates or causing a complete loss of service.

Repeatability

One of the difficulties with weighting criteria is dealing with non-rational influencing factors. If the stakeholders in a weighting workshop had a similar bad experience, you might expect them to apply more weight to criteria appropriate to that experience. On the face of it, the outcome is mutually agreed, buy-in has been achieved, and the outcome is sound. However, it could be that the shared experience (or effect) had a cause that is not appropriate to the problem. In the event that the stakeholders had different experiences, would the relative weights be the same? Does this mean that the influences are non-rational? To determine the answer to these questions, the weighting process needs to discover (and record) why the opinions are held. By doing so, the weighting process is more likely to yield a result that would be repeatable on a different day.

There are various techniques that can be used to help people understand why they hold certain views. A technique called 'Five Whys'[6] can help, as long as the method is not used in a confrontational manner.

Level	Description	Value
Very significant impact	Cost [± 100%] Time [± 100%]	8
	Quality (complete loss of service/project − or − expectations exceeded)	
…	…	7-2
Very low impact	Cost [0]	1
	Cost [0]	
	Quality (service as expected)	

Table 3. Objective weighting descriptions

Structured criteria weighting

This method is a practical mechanism that has been developed by QinetiQ Commerce Decisions primarily to address procurement decisions.

The maths is non-complex, and the resulting weights are auditable and defensible.

The method examines each criterion using four factors:

– impact

– distinctive capability

– certainty

– immediacy

Each factor is scored using an appropriate scale, justified by the scorer (or stakeholder). The four scores are combined to give a weight for the criterion.

The four factors act as a framework around which the stakeholders can discuss and record the importance of the criteria.

Here are two examples of justifications recorded (the words in capitals are the score given for each factor):

Weight 100:

– **Impact (VITAL)**: The successful seamless delivery of this service area is vital to the Authority. Once in operation there is no fall-back position.

– **Distinctive capability (RARE)**: To our knowledge industry has not had to deliver such a capability before.

– **Certainty (FULL INFO)**: The Authority has been able to clearly articulate and detail the requirements in such a manner as to allow the tenderers to provide a comprehensive response.

– **Immediacy (IMMEDIATE)**: This capability needs to be in operation within one calendar month after the contract is let and there is no time for corrective action.

Weight 28:

– **Impact (IMPORTANT)**: The delivery of this service area is important to the Authority. However, in the event that the service should falter, the Authority is able to offer assistance from resource pool X.

– **Distinctive capability (UNIVERSAL)**: All industry has knowledge of the service, as similar services have been outsourced for many years through various Authorities.

– **Certainty (PARTIAL)**: There are gaps in the available information and the tenderer will need to make assumptions that may later turn out to be incorrect.

– **Immediacy (MEDIUM)**: The capability needs to be in place by mid-contract.

If a challenge was made to the weight of a criterion (which sometimes happens with non-involved senior stakeholders), a justification will be available to defend it. If you wanted to change a weight it would be reasonable to ask for the justification to be changed also.

This method is used as part of the Structured Criteria Development process developed by QinetiQ Commerce Decisions. The four factors are tailored to procurement. For different problem areas (such as option analysis, or operational effectiveness analysis) a different framework of view points is used.

QinetiQ Commerce Decisions has had very positive feedback from many clients using the Structured Criteria Weighting technique, primarily due to how effectively it engages stakeholders and produces defensible weighting schemes.

Practicalities

The weights that we use must be defendable and dependable. The worst situation is to produce a weighting model that (when populated with tender returns) yields a result that no one has any faith in. The Structured Criteria Weighting method has been designed particularly to deal with this problem.

Garbage in – garbage out. No process will provide a good solution unless the right data (in this case human intelligence) is applied considerately.

The use of an experienced independent facilitator to challenge inputs can greatly increase the quality of the weights model.

Once a model has been established, it is important to test it. There are a number of mechanisms used for testing models. The most appropriate one will depend on the complexity of the model. However, it is worth bearing in mind that it is easy, but time-consuming and often not useful, to provide a mathematical analysis of a model. Running simulations is also possible. However, to understand the real implications of the model, the analysis must cover the relationship between the details of the subjects and the weights. The analysis should give feedback commentary such as:

"The model implies that you are willing to trade small elements of subject matter expertise in area A for more content in areas C, D and E."

This is best achieved by using an experienced practitioner to provide feedback that relates to the intellectual side of the model rather than the purely mathematical.

Conclusion

There are many methods and variations that can be used to determine weights for criteria. The challenge is to choose a method that is appropriate to the task. As such, any weighting practitioner must have knowledge of different methods, and the pros, cons, risks and benefits of each.

Failing to consider weighting properly can lead to selection of the wrong supplier and ultimately failure of the project. Using the right techniques it is possible to reflect all stakeholders' needs, defend the decisions, and emphasise the criteria that are going to make a difference to the project outcome.

The end result of the selection process should be an anti-climatic decision, where the result appears obvious. Creating the 'right' weights for the criteria is a fundamental step required to achieve this.

References

1 [2007] EWCA Civ 1522 (21 December 2007)

2 In Case C-532/06 – Lianakis and Others v. Dimos Alexandroupolis and Others

3 Salo, A. A. and Hämäläinen, R. P., 'On the measurement of preferences in the analytic hierarchy process', Res. Rep. A47, Helsinki University of Technology, Systems Analysis Laboratory, 1993

4 Ma, D. and Zheng, X., `9/9±9/1 scale method of AHP', Proc. 2nd Int. Symp. on the AHP, Pittsburgh, PA, 1991, Vol. I, pp. 197±202.

5 Saaty, T. L., The Analytic Hierarchy Process, New York: McGraw-Hill, 1980.

6 Developed by Sakichi Toyoda founder of Toyota Industries Co. Ltd. Originally created to help solve systems problems.

About Richard Tottman

Richard Tottman is a principal consultant at Commerce Decisions. He has facilitated weighting workshops for numerous multi-million pound programmes. He is a practitioner of the Commerce Decisions Structured Criteria Development process.

Richard also co-wrote the early versions of AWARD. Before 2001 Richard was a requirements management consultant and software development manager, working for a number of organisations including BAE Systems and Thales.

About Commerce Decisions

Commerce Decisions has been supporting strategic, high-risk procurements globally since 2001, and is at the forefront of best practice procurement. With a unique focus on complex evaluation, we have unrivalled experience in supplier evaluation and are a trusted provider of procurement services to the public and private sectors.

We deliver a robust and defensible procurement process to our clients, proven time and time again across many sectors including construction, transport, education, health, defence and facilities management procurements – to date, we have supported over 13,000 strategic projects, collectively worth over $400billion. This enviable experience and in-depth knowledge has enabled us to develop proven methodologies, supporting clients to deliver the best possible outcome on strategic and complex procurement projects.

Headquartered in Oxfordshire, UK, and with offices in Canberra, Australia, and Ottawa, Canada, Commerce Decisions provides software and services to support the procurement and post contract review processes for both buyers and suppliers. For buyers we improve the efficiency and effectiveness of the evaluation process to make the best buying decision based on all the relevant criteria, underpinned by our AWARD® software. For bidders we improve the quality and timeliness of proposals to best meet the needs of the potential buyer and thereby give them the best chance of securing the contract, underpinned by our ADVANCE™ software.

About QinetiQ

QinetiQ is a leading international provider of technology-based services and solutions to the defence, security and related markets. We develop and deliver services and solutions for government organisations, predominantly in the UK and US, including defence departments, intelligence services and security agencies.

In addition, we provide technology insertion and consultancy services to commercial and industrial customers around the world.