Weigh the scales

by Alex Zhykharyev, Ph.D.

The constraints of measurement theory on the proposal evaluation and selection process

According to its common usage within the public procurement world in Canada,[1] "A Request for Proposal (RFP), while generally used for requirements of $25,000 or more is often employed for purchases where the selection of a supplier cannot be made solely on the basis of the lowest price. An RFP is used to procure the most cost-effective solution based upon evaluation criteria identified in the RFP."

In response to a RFP, suppliers are usually expected to submit detailed proposals. The choice of the supplier is the result of a proposal evaluation and selection process.

Let's consider that the proposal evaluation and selection process includes the following consecutive steps:

- **Technical evaluation step.** The purpose of this step is to measure the technical value of each proposal. Evaluators are supposed to score each proposal against a set of predefined criteria. The scores of each criterion are totalled to give the final score for each proposal, which represents the measurement of its technical value.
- **Cost evaluation step.**
- **Selection step.** This step uses both the technical evaluation scores and cost evaluation to show the most cost-effective solution.

At the technical evaluation step, the scoring of proposals produces results based on different scales or measurement levels. How the scores for criteria and their weights are defined determines the type of scale – either interval or ratio – being used.

According to the measurement theory, the type of measurement scale of the total scores obtained at the technical evaluation step imposes constraints on which selection methods can be used at the selection step and, conversely, a given method used at the selection step imposes constraints on the type of measurement scale of the total scores. These constraints must be taken into account if our selection method is not to be based on the results of meaningless statements.

Sarle[2] describes a commonplace meaningless statement as "the claim by the weatherman on a local TV station that it was twice as warm today as yesterday because it was 40 degrees Fahrenheit today but only 20 degrees yesterday. This statement is meaningless because one measurement (40) is twice the other measurement (20) only in certain arbitrary scales of measurement, such as Fahrenheit. The relationship "twice-as" applies only to the numbers, not the attribute being measured (temperature)." An evaluation and selection process that uses the results of meaningless statements to determine the outcome and choice of solution may find that the chosen solution is not necessarily the most cost-effective solution. Moreover, the final choice could be the least cost-effective solution. And clearly it is not the quality of the final choice, which is supposed to be the most cost-effective solution.

To illustrate the effects of making decisions based on the measurement results of differing scales, let us look at what type of scales are needed for two selection methods described in the *New Buyers' Guide* published by Contracts Canada:[3]

- **Highest technical score**
  The highest evaluated proposal wins, regardless of cost.
- **Best value method (BVM)**
  In this method, supplier selection will be based on the bidder offering the lowest responsive cost-per-point proposal, determined by dividing the bid price by the total points achieved in the technical evaluation of the bidder's proposal. The firm with the lowest cost per point is selected.

As mentioned earlier, the total scores produced at the technical evaluation step are interval and ratio scales. The highest technical score method can use the total scores of either interval or ratio scales. The best value method (BVM) can use only the total scores of ratio scales.

Now imagine that we want to select the better of option A or option B using the BVM. Assume that the technical value (TV) of the options is expressed in terms of temperature,

where higher temperatures have higher value. Let the cost of option A be equal to $30,000 and the cost of option B be equal to $10,000. Since the cost of option A is three times as great as the cost of option B, option A can be chosen by the BVM only if its TV is more than three times greater than the TV of option B. Let us use two different temperature scales, Fahrenheit and Celsius, to compare the TV of our options.

First measure the TV of the options in terms of Fahrenheit scale. Let options A and B have the TVs 80 and 40 degrees Fahrenheit, respectively. Comparing the figures 80 and 40, we conclude that option A has a TV twice as great as the TV of option B. Thus the BVM must select option B.

Now use the Celsius scale. The TV of option A becomes 27 and the TV of option B becomes 4. Option A is now portrayed to have a TV of nearly seven times the TV of option B. Consequently, we must come to the conclusion that option A must be chosen on the basis of the BVM.

How, by using either the Fahrenheit or Celsius scales, which both measure the same attribute (temperature) to determine best value, did we produce opposite results? Both scales are interval measurement scales where the ratio between any two measurements has no significance. Interval scales have no fixed zero that represents a zero quantity of the quality being measured, therefore, ratios between any two measurements on the scale cannot be meaningfully constructed. In general, for any given interval scale, one can always build (or find, as in our example) another interval scale that produces the opposite result. Both scales, the given and new, can be used to measure TV.

So, when TV measurements use an interval scale, the choice made between any two given options produced by the BVM is meaningless. And if measurement theory is ignored and you choose to use an interval scale to determine the solution under the BVM, then the result will be arbitrary – possibly not the most cost effective.

One might assume, that in practice, public procurement selection processes apply only ratio scales for the BVM, and that there is no cause for concern. Unfortunately, most of the methods applied to "assign values out of 100" for technical evaluation (TV) produce total points in interval scales. This includes a decision support software product (DSSP) used by many governments and departments to support public sector procurement decisions.

DSSP uses the approach of multi attribute utility theory (MAUT) to obtain total points for proposal technical evaluation. MAUT can guarantee only interval scales for total points. As a result, best overall value (BOV) and many other "advanced" methods of this DSSP are based on meaningless statements. When setting up the evaluation process, buyers should consider carefully the type of scale they or their DSS will use to determine best value. As shown in the Fahrenheit-Celsius example above, the use of interval scales to determine best overall value can lead to quite arbitrary conclusions, which may not, in fact, provide best value.

comstock.com

Dr. Alex Zhykharyev completed his post doctoral study, Applied Decision Theory and Operations Research, at the National Technical University of the Ukraine. A freelance software developer, he is based in Toronto, Ontario and can be reached at azhykh@estart.com or (416) 259-8866.

[1] Glossary of Procurement Terminology. *Summit Magazine.* Retrieved August 31, 2006 from http://www.summitconnects.com/Tool_Kit/glossary_N_Z.htm#R

[2] Sarle, W.S. (1996) *Measurement theory. Frequently asked questions.* SAS Institue Inc. Retrieved August 31, 2006 from ftp://ftp.sas.com/pub/neural/measurement.faq

[3] *New Buyers' Guide. Chapter 10 – Evaluating tenders and proposals,* (2004), Contracts Canada. Retrieved August 31, 2006 from http://contracts canada.gc.ca/en/chap10-e.htm