

THE NASH & CIBINIC REPORT®

*government contract analysis and advice monthly
from professors ralph c. nash and john cibinic*

THOMSON
WEST

Author: Ralph C. Nash, Professor Emeritus of Law, The George Washington University
Contributing Author: Vernon J. Edwards

VOLUME 20 • NUMBER 2
FEBRUARY 2006

INSIDE:

COMPETITION & AWARD

Scoring or rating in source
selection: a continuing source of
confusion ----- ¶ 7

COSTS & PRICING

Postscript IV: the "jury verdict"
approach ----- ¶ 9

Cost Principles changes chart:
FAR 2005 edition ----- ¶ 10

PROCUREMENT MANAGEMENT

Obligating funds under the
Economy Act: a clarification ----- ¶ 8

The Nash and Cibinic Report (ISSN 0891-9291) is published monthly and copyrighted © 2006 by Thomson/West • Published by Thomson/West / 610 Opperman Drive, P.O. Box 64526 / St. Paul, MN 55164-0526 • <http://www.west.thomson.com> • Periodicals postage paid at St. Paul, MN • Postmaster: Send address changes to The Nash & Cibinic Report / PO Box 64526, St. Paul, MN 55164-0526.

This publication was created to provide you with accurate and authoritative information concerning the subject matter covered; however, this publication was not necessarily prepared by persons licensed to practice law in a particular jurisdiction. The publisher is not engaged in rendering legal or other professional advice, and this publication is not a substitute for the advice of an attorney. If you require legal or other expert advice, you should seek the services of a competent attorney or other professional. Reproduction, storage in a retrieval system, or transmission of this publication or any portion of it in any form or by any means, electronic, mechanical, photocopy, xerography, facsimile, recording or otherwise, without the written permission of Thomson/West is prohibited. For authorization to photocopy, please contact the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923, (978)750-8400; fax (978)646-8600 or West's Copyright Services at 610 Opperman Drive, Eagan, MN 55123, fax (651)687-7551.

40399061

Competition & Award

¶ 7 • SCORING OR RATING IN SOURCE SELECTION: A Continuing Source Of Confusion • Two recent protest decisions show that some Contracting Officers still do not understand the difference between evaluating proposals and scoring or rating them and do not understand the proper role of scores or ratings in

Notes

contractor selection processes. In *Beta Analytics International, Inc. v. U.S.*, 67 Fed. Cl. 384 (2005), *declaratory relief ordered*, 2005 WL 3150612 (Fed. Cl. Nov 23, 2005), 47 GC ¶ 524, the U.S. Court of Federal Claims decided a postaward bid protest in favor of the plaintiff because the source selection official relied on unsupported average scores in making her decision. In *YORK Building Services, Inc.*, Comp. Gen. Dec. B-296948.2, 2005 CPD ¶ 202, 47 GC ¶ 537, the Comptroller General sustained the protest because the source selection official relied on unsupported total scores to make a decision that was inconsistent with the terms of the Request for Proposals.

Needless Complexity, Careless Execution, Ensuing Confusion, And Avoidable Error

In *Beta Analytics*, the Naval Sea Systems Command conducted a procurement for scientific, engineering, and technical assistance services for the Defense Advanced Research Projects Agency. Beta Analytics (BAI) was the incumbent contractor and lost the source selection to Maden Tech Consulting, Inc. As explained by the court, the Navy used a numerical/adjectival scoring scheme and, eschewing the virtues of simplicity, the following evaluation factors and point distributions:

- I. Technical Proposal
 - A. Program Plan (45 points)
 1. Overall Proposal (20 points)
 2. Risks Associated With Contract Performance (5 points)
 3. Measurement of Provided Services (5 points)
 4. Contract Transition (5 points)
 5. Staffing Plan (5 points)
 6. Corporate Support (5 points)
 - B. Experience (30 points)
 1. Counterintelligence (5 points)
 2. International Security (5 points)
 3. Communications Security (5 points)
 4. Information Security (5 points)
 5. Industrial Security (5 points)
 6. Program Security (5 points)
 - C. Personnel (25 points)
- II. Past Performance (Adjectival: Poor to Excellent)
- III. Cost/Price (Not scored)

According to the court, the scoring sheets used by the evaluators for the “Personnel” factor and overall “Program Plan” factor contained “fairly detailed scoring criteria.” But the guidance for the “Program Plan” and “Experience” subfactors was limited to the following two sentences:

Maximum Points: Meets every aspect of the requirement and exceeds the requirement in some areas with no exceptions or weaknesses. Minimum Points: Does not meet the requirement and did not demonstrate ability to meet the requirement.

Under that guidance each evaluator was to assign whatever score the evaluator thought appropriate.

The evaluation panel chairman was to calculate each offeror’s average score from the scores given by the evaluators. Those numbers were then to be provided to the source

Notes

selection authority accompanied by supporting “narrative.” The scores assigned by the Technical Proposal evaluators to BAI and Maden are displayed in the table below.

Technical Proposal Evaluation Scores								
	BAI				Maden			
Evaluator	Plan	Exp.	Pers.	Total	Plan	Exp.	Pers.	Total
Aronson	43	30	24	97	44	26	23	93
Bailey	*34	27	23	*84	40	16	22	78
Gartrell	33	22	15	70	43	25	25	93
Average	*36.66	26.33	20.66	*84	42.33	22.33	23.33	88

(The asterisks signify that evaluator Bailey did not score BAI on the “Risk” and “Contract Transition” subfactors, because he did not think that they applied to the incumbent contractor.)

The court found gaps and discrepancies in the Navy’s documentation of the evaluation process, its findings, the scoring, and the calculations of average scores and found that the “narrative” that accompanied the scores did not adequately explain the differences among the proposals. Illustrative of the kinds of problems that troubled the court were the scores that evaluator Gartrell gave to BAI and to Maden Tech for the “Personnel” factor. She gave 15 points to BAI, just above the minimum of 14 and significantly lower than the scores given to BAI by the other evaluators—23 and 24 points, but she gave Maden Tech a perfect score of 25. According to the court: “In awarding BAI a 15, the only notes made by Gartrell appear to read ‘one of the key personnel proposed lack [sic] the educational requirement outlined in the Solicitation,’ and ‘[XX (position A) XX] does not detail [Department of Defense] related experience.’” On the other hand, Gartrell gave Maden Tech a perfect score “[w]ith no explanatory notes.”

In its lengthy description of the Navy’s best value tradeoff memorandum, the court said:

At the end of the January 26, 2004 memorandum, the value and price tradeoffs are performed, presumably by the signatories—evaluation board chairman McClure and Contracting Officer Kennedy. Concerning the Technical Proposals, only the final average technical evaluation scores are mentioned—Maden Tech’s 88, BAI’s 84 [XXXX]’s [XX]. *See id.* at 1003. But sandwiched in-between the table on the first page and the tradeoff’s on the last is a summary discussion of each proposal. This discussion breaks each offeror’s score down to the factor and sub-factor level. [Footnote omitted.] The narrative that accompanies each factor and sub-factor is, for the most part, made up of either paraphrasing or verbatim excerpts from the respective offeror’s proposal, with an occasional description of features such as charts.

* * *

The value and price tradeoffs, as was noted above, were based on the total score for each offeror’s Technical Proposal, and made no mention of any differences among the specific features of those proposals. Following the Solicitation, *see Admin. R.* at 175, 186, the memorandum stated that “[t]he technical evaluation is the most important factor and price is the least important factor.” *Id.* at 1003.... Comparing Maden with [XXXX], it was noted that [XXXX]’s technical score was “in the marginal range compared to Maden Tech’s score, which is in the high range for success,” and was determined that Maden Tech’s “[XX]% higher score” in the technical area was “worth” the “12% higher price.” *Id.* The memorandum concluded: “Maden Tech’s proposal is considered to be the Best Value to the government and is recommended for the award. *Id.*”

In the end, the court found that the scores given to BAI for certain “Program Plan” subfactors and for “Personnel” were based on evaluator error or faulty procedure. The court also found that the score given to Maden Tech for “Personnel” was based on evalu-

Notes

ator error and that, in general, the technical scores given to Maden Tech were based on faulty procedure. The court concluded that the proposal scoring and the decision were arbitrary and capricious. It issued a permanent injunction against the exercise of any options in the Navy's contract with Maden Tech and ordered a new procurement to be completed before April 9, 2006.

In the Comptroller General's *YORK* decision, which involved a procurement of janitorial services by the Department of Agriculture, the agency evaluated offerors' "technical" approaches (for janitorial services?!), management plans, past performance, and prices. The RFP said that the three "technical" factors were more important than price. The Comptroller General sustained the protest based on the finding that the evaluators scored the proposals in a manner that was inconsistent with the RFP's statement of the relative importance of the evaluation factors:

The record confirms that, in evaluating proposals, the evaluators treated the technical approach and management plan evaluation factors as having equal weight by allowing a maximum of 35 points to each. AR, Tab D, Standards for Evaluation of Proposals, at 4. The weighting was contrary to the terms of the RFP, which indicated that technical approach was the most important evaluation factor....The [Source Selection Official] appears to have then relied on those point scores in a mechanical fashion in concluding that the proposals of [the protester] and [the awardee] were technically equal and in awarding to [the awardee] on the basis of its lowest price. Point scores are merely a guide to intelligent decision-making and their proximity does not, itself, establish technical equality. See *Ogden Support Servs. Inc.*, B-270012, B-270012.2, Mar. 19, 1996, 96-1 CPD ¶ 166 at 6.

The Comptroller General noted that the source selection official's comments "lacked any indication that she considered, or was aware of, the substance of the technical proposals or any evaluation consistent with the criteria in the RFP." The Comptroller General recommended that the agency reconsider its evaluation scheme, conduct a new procurement if necessary or reevaluate the proposals and make a new source selection decision, and reimburse the protester for the costs of filing and pursuing the protest.

In both of these procurements, the agencies used numerical scoring, and the likely reactions to these decisions will be that they confirm the dangers of numerical scoring. But what the decisions really confirm is that any scoring or rating scheme is dangerous in the hands of incompetent people, no matter what device is used—numbers, adjectives, colors, stars, or others.

The Nature And Process Of Proposal Evaluation

Strictly speaking, agencies do not evaluate proposals, they evaluate the offerors and offers (promises) described in proposals. Evaluation factors are *attributes* of an offeror and of what it promises to do or deliver—features, qualities, and characteristics that a buyer wants or does not want. The relationship between attributes and value is cause to effect. Offerors and offers have value to the extent that they have attributes that the buyer wants and do not have attributes that the buyer does not want, and evaluation should aim at making sound value judgments about offerors and offers based on their attributes.

It is essential that everyone involved in contractor selection understands the distinction between *evaluating* offerors and their offers and *scoring* or *rating* them. Evaluation is the process of determining the relative value of a thing. Scoring or rating is the use of words or symbols to express evaluation findings in simple terms. The Federal Acquisition Regulation requires agencies to evaluate offerors and their offers, but it does not require that they score or rate them. FAR 15.305(a) is misleading in saying that evaluations "may be conducted using any rating method or combination of methods, including color or adjective ratings, numerical weights, and ordinal rankings," because rating, in the sense of the assignment of adjectives or

Notes

symbols, is not a method of conducting an evaluation, it is a method of expressing the results of an evaluation.

The first step in evaluation is to factually and judgmentally determine what attributes are present or absent in an offeror and in its promises, and in what degree. The second step is to make judgments about the relative value of each offeror and its offer in comparison with other offerors and their offers. Thus, evaluation consists of two steps: (1) identifying and describing the attributes of each offeror and its offer, and (2) determining and describing the relative merits of each in comparison with others based on the differences in their attributes. The second step is the ranking and decision step, and it is in this step that scores or ratings are both useful and potentially misleading.

Scores or ratings (the FAR uses the term “rating”) summarize complex information about the attributes of offerors and offers, making it easier to see the big picture at a glance. They do this by converting detailed information about a proposal into a short-hand expression—usually a number or an adjective—that represents some abstract measure of value. On a scale of 0 to 100 points, 100 being best, a score of 80 summarizes the merits of an offeror and its offer on an abstract scale of value, but the score does not tell you anything about the sources and nature of that value. Thus, the big picture comes at the cost of a considerable loss of information. Given such loss of information, why score? The answer is that scoring responds to the demand: *Cut to the chase! How good are the proposals?*

What everyone must understand about scores and ratings is that since they summarize detailed factual findings and judgments about the presence or absence of attributes, the technique of *first* assigning scores or ratings and *then* preparing documentary explanation or justification for the scores is unsound. The evaluators must *first* determine and document what attributes are present in the offeror and its offer, and only *then* assign scores or ratings that reflect those attributes. The reliability of scores or ratings as information depends on how they are assigned; their usefulness as information depends on how well the decisionmaker understands what they are supposed to mean.

The Uses And Limitations Of Scores And Ratings

To better understand the issues in scoring or rating system development, consider the following example. Suppose that an agency is going to solicit proposals for an environmental remediation project of moderate difficulty and plans to evaluate offerors and their offers the basis of the following five factors and numerical point distributions:

Evaluation Factor	Maximum Points
Scope of Experience	30
Soundness of Technical Approach	25
Soundness of Management Approach	25
Quality of Key Personnel	10
Quality of Past Performance	10
Price	Not to be scored

The RFP tells offerors that the agency will evaluate the scope of each offeror’s experience on the basis of the number of remediation projects that it has undertaken within the last three years that were of similar size to the project described in the RFP and that are either completed or at least half completed and on the complexity and difficulty of those projects. The RFP defines “experience” as follows:

Experience is the opportunity to learn through performance of projects similar to the one described in this RFP. Its assessment and value to the Government are matters of

Notes

its *scope*, which means the number and variety of tasks performed, the variety of conditions under which the work was performed, and the difficulty of performance. The greater the number of projects undertaken, the greater the variety of tasks performed, the greater the variety of conditions in which the tasks were performed, the more difficult the tasks, and the more complete the projects, the greater the scope of experience and its value to the Government.

Assume that the agency gives the evaluators the following instructions about scoring experience:

After documenting the facts and your judgments about an offeror's experience, assign a score in the range of 1 to 10 points if the offeror undertook one or two projects of similar size to the one described in the RFP and that are either completed or at least half completed. Assign a score in the range of 11 to 20 points if the offeror undertook three or four similar projects. Assign a score of 21 to 30 points if the offeror undertook five or more similar projects. Based on your knowledge and experience, assign a specific score within those ranges on the basis of the variety, complexity, difficulty, and state of completion of the projects.

This is a *scoring rule*. (Some agencies refer to such a rule as an *evaluation standard*.)

The scoring rule leaves considerable room for judgment (and inconsistency) in the assignment of scores. While it should be clear that an offeror that receives 30 points has better experience than an offeror that receives 15 points, the scores do not necessarily indicate how much better that experience is. The scoring rule does not define a unit of experience and the numbers 30 and 15 are not quantities of experience units. So what does a number of points signify?

The scoring rule establishes *value functions* for sets of attributes—*scope* of experience, *soundness* of approach, *quality* of personnel, and *quality* of past performance. A score should reflect an evaluator's judgment as to the abstract *value* of the scope of an offeror's *experience*, expressed in numerical form. The higher the number the better the adjudged value of the offeror's experience. A score of 30 points indicates twice as much adjudged value as a score of 15 points, but not necessarily twice as much experience. Think of *value* as a common denominator, an attribute which subsumes all of the others. Twenty points of one factor are worth 20 points of another.

Now suppose that our agency receives two proposals. After reading them, the evaluators document the following facts and judgments:

Offeror A has undertaken six similar projects within the past three years. Two are completed, one is half completed, and two are three-quarters completed. The two completed projects encompassed sites of between one-half to one square acre. Both were of moderate complexity and difficulty. The half-completed project encompasses a site of one and one-half square acres. It, too, is of moderate difficulty. The three-quarters completed projects both encompassed sites of less than one-half acre and were simple and not especially difficult. All projects were performed in semi-rural areas with good road access to the sites. All of the projects were performed in southern California, in good weather conditions.

Offeror B has undertaken eight similar projects within the past three years. Four are completed, two are three-quarters completed, and two are half completed. All of the projects encompassed sites of less than one square acre. Six of the projects were performed in urban or suburban areas and entailed compliance with a complex set of local and regional ordinances in addition to federal laws and regulations and coordination with local and regional authorities. They were of moderate to high complexity and difficulty. The other two were performed in semi-rural areas with poor road access. They were both of moderate difficulty.

Notes

Three of the projects were performed in the Boise, Idaho area, with good to moderate weather. Three were performed in western Washington, where the weather was often rainy. Two were performed near Fairbanks, Alaska, under difficult weather conditions.

Based on those facts and judgments, the three evaluators assigned experience scores as follows:

Evaluator	Offeror A	Offeror B
Nash	25	29
Cibinic	21	25
Edwards	23	21
Average Score	23	25

After discussion among themselves, the evaluators agreed on final consensus scores of 22 for Offeror A and 26 for Offeror B (different from the average scores). The evaluators used similar guidance and followed similar procedures to score the offerors on the other factors. Upon completion of the evaluation, the evaluators presented the Source Selection Authority (SSA) with the following summary information:

Evaluation Factor	Offeror A	Offeror B	Margin/Advantage
Scope of Experience	22	26	4 points/B
Soundness of Technical Approach	20	18	2 points/A
Soundness of Management Approach	20	15	5 points/A
Quality of Key Personnel	10	6	4 points/A
Quality of Past Performance	7	9	2 points/B
Total Nonprice Score	79	74	5 points/A
Price	\$5,981,000	\$5,750,000	\$231,000/B

This summary information enables the SSA to quickly determine where the offerors stand in the eyes of the evaluators, but the scores do not tell the SSA the reasons for that standing or enable the SSA to make an intelligent tradeoff decision. Offeror A has a five point advantage over Offeror B in terms of the nonprice factors, but the scores do not tell the SSA the content of Offeror A’s advantage, or whether it is worth \$231,000. To know that, the SSA must refer to the evaluators’ documentation of their factual findings and judgments.

Scoring Or Rating With Adjectives And Colors

Suppose that instead of numbers the agency used adjectival scoring with the following scale:

- Excellent
- Good
- Adequate
- Inadequate

Assume that “Excellent” experience is the equivalent of 21 to 30 points on the numerical scale, “Very Good” experience is the equivalent of 11 to 20 points, and “Adequate” experience is the equivalent of 1 to 10 points. Assume further that the agency provided the evaluators with scoring guidance similar to that provided in the numerical scoring example.

Notes

The adjectival scale is ordinal. It tells you that “Excellent” is better than “Very Good,” but it does not reveal how much better. Also, two proposals might be “Excellent,” but one might be more “Excellent” than the other. Here is the same evaluation outcome discussed above expressed in adjectival terms:

Evaluation Factor	Offeror A	Offeror B	Advantage
Experience	Excellent	Excellent	B
Soundness of Technical Approach	Excellent	Excellent	A
Soundness of Management Approach	Excellent	Good	A
Quality of Key Personnel	Excellent	Adequate	A
Quality of Past Performance	Adequate	Excellent	B
Summary Nonprice Score	Good	Good	A
Price	\$5,981,000	\$5,750,000	\$231,000/B

The adjectival scores could be supplemented with + and – signs to permit more differentiation in an attempt to create an interval scale, in which case the outcome might look like this:

Evaluation Factor	Offeror A	Offeror B	Advantage
Experience	Excellent	Excellent+	B
Soundness of Technical Approach	Excellent+	Excellent	A
Soundness of Management Approach	Excellent+	Good–	A
Quality of Key Personnel	Excellent++	Adequate	A
Quality of Past Performance	Adequate	Excellent	B
Summary Nonprice Score	Good+	Good	A
Price	\$5,981,000	\$5,750,000	\$231,000/B

The plus signs give the reader a sense of the size of the differences between the proposals.

Is Scoring Or Rating A Good Idea?

In light of the limitations inherent in all scoring and rating methods, is scoring or rating a good idea? Sure. Scores or ratings are useful to an SSA who wants the “big picture” so the SSA can decide what questions to ask the evaluators about specific findings and judgments, questions such as: *Specifically, how were the offerors’ technical and management approaches different? What good things did Offeror A have in its approaches that Offeror B did not, or what bad things did Offeror B have in its approaches that A did not?* And it is understandable that an SSA would say: *Cut to the chase! What’s the bottom line? Sum it up!*

While we think that the Court of Federal Claims made the right decision in *Beta Analytics*, we disagree with the court in one respect. The court’s decision suggests that it would have been okay for the Navy to have chosen the contractor on the basis of a comparison of average numerical scores had the scores been rationally founded. While such comparisons may be acceptable to the court, we think that they are a fundamentally unsound practice. Scores or ratings may be helpful in contractor selection, but they are not sound bases for decisionmaking. No decisionmaker should ever base tradeoffs or source selection decisions on scores or ratings. If I were a CO today and were using a scoring or rating scheme, I would not describe it in an RFP or mention scores or ratings in a decision document. I would insist that decision documents explain tradeoffs and the decision rationale strictly in terms of specific attribute differences—good things and bad things that contribute or detract value. *VGB*