



Decision

Matter of: Hughes Training, Inc.
File: B-256426.4
Date: January 26, 1995

Thomas E. Hill, Esq., Haynes and Boone, L.L.P., for the protester.
Donald J. Walsh, Esq., Wright, Constable & Skeen, for AAI Corporation, an interested party.
Curtis D. Elton, Esq., and Gregory H. Petkoff, Esq., Department of the Air Force, for the agency.
M. Penny Ahearn, Esq., David A. Ashen, Esq., and John M. Melody, Esq., Office of the General Counsel, GAO, participated in the preparation of the decision.

DIGEST

1. Protest against evaluation of protester's computer software capabilities as presenting high risk is denied where protester failed to furnish requested historical data regarding validity of its software estimating methodology and the historical data which it did submit did not demonstrate the reliability of the firm's software development estimates.
2. Agency's approach to estimating most probable cost (MPC) of required computer software development effort for aircraft maintenance trainer is unobjectionable where the agency used a commercial software estimating program to arrive at offeror-unique MPC based on an adjusted cost model, and the record provides no basis to question the validity of either the information or the approach used.

DECISION

Hughes Training, Inc. protests the Department of the Air Force's award of a contract to AAI Corporation, under request for proposals (RFP) No. F42630-93-R-2037, for a maintenance training set prototype for the Joint Surveillance Target Attack Radar Systems (JSTARS E-8C).

*The decision issued on January 26, 1995, contained proprietary information and was subject to a General Accounting Office protective order. This version of the decision has been redacted. Deletions are indicated by "[deleted]."

Hughes challenges the agency's evaluation of the technical and cost proposals.

We deny the protest.

The RFP contemplated award of a cost-plus-award-fee contract for the development of a prototype maintenance training set for the JSTARS E-8C, with options for one production unit, logistics support for each trainer, and for a training systems support center. The solicitation listed as evaluation factors: (1) management/logistics support and (2) technical/test, which were of equal importance, and (3) cost, which was of less importance. The management/logistics support and technical/test factors included seven subfactors each, listed in descending order of importance. At issue here is the evaluation of software capabilities, the second subfactor under the technical/test factor. The solicitation provided for each of the non-cost factors and subfactors to be given a color/adjectival rating, a proposal risk rating, and a performance risk rating. With respect to cost, the RFP provided for the evaluation of cost reasonableness, cost realism, and cost completeness. The RFP, as amended, advised offerors that cost realism would be evaluated using the government's estimate of most probable cost (GEMPC) for each offer.

Four proposals were received in response to the solicitation. After conducting discussions and obtaining best and final offers (BAFO), the Air Force made an initial award to AAI. In response to Hughes's subsequent protest to our Office against the evaluation of technical and cost proposals and the conduct of discussions, the agency amended the solicitation primarily to provide for cost evaluation on the basis of the GEMPC and reopened discussions.

After discussions were reopened, additional clarification requests (CR) were issued to the offerors. The CRs issued to Hughes included requests for information concerning the validity of the firm's software estimate, measured in lines

¹The color ratings were blue-exceptional, green-acceptable, and yellow-marginal. The risk ratings were high, moderate, and low. The color/adjectival rating was to assess how well the offeror's proposal met the evaluation standards and solicitation requirements. Proposal risk was to assess the risk associated with the offeror's proposed approach as it related to accomplishing the requirements of the solicitation. Performance risk was to assess the probability of the offeror successfully accomplishing the proposed effort based on the offeror's demonstrated present and past performance.

of code (LOC). An additional round of BAFOs then was requested.

Based upon its evaluation of the second BAFOs, the source selection authority determined that AAI's proposal offered the best overall value. AAI's proposal was rated green and low risk in all areas; the proposal was evaluated as having the lowest GEMPC of \$13,623,488, excluding award fee (with a proposed cost of \$10,683,070). In addition, AAI's offer to accelerate the performance schedule by 3 months was considered a substantial benefit to the government. Hughes's proposal was rated green and low risk in all areas except for the software capabilities subfactor, where it received a high proposal risk rating due to the evaluated lack of validation for its software estimate. As a result, although Hughes's proposed cost was \$9,480,803, the agency calculated its GEMPC as \$15,659,004 (excluding award fee) for the proposal, the highest of any offeror. In essence, the agency determined that Hughes had underestimated the software development effort to such an extent that there would be a high risk that Hughes could not maintain its schedule and cost projections. Upon learning of the agency's selection of AAI for a second time, Hughes filed this protest with our Office.

Hughes primarily challenges the agency's evaluation of the firm's software capabilities and its calculation of the most probable cost of Hughes's proposal. Based upon our review of the record, we find no basis to question the Air Force's determination that Hughes significantly underestimated both the required software development effort and the overall cost of performing. We discuss the protester's principal arguments below.

TECHNICAL EVALUATION-SOFTWARE CAPABILITIES

Background

The sole area of difference in the technical evaluation between the two offerors was under the software capabilities factor, where Hughes's proposal was rated high risk while AAI's was rated low risk. The RFP instructed offerors to furnish detailed information concerning software capabilities, including such items as: (1) "a complete set of internal company/division software management standards, procedures, methods, operating instructions or other forms of internal guidance and direction," with descriptions of "software estimating: size, manpower, schedule, distribution of manpower over the schedule and cost"; (2) "examples of models and methods of application to estimate software size, and the associated manpower effort, cost, schedule, and distribution of manpower over the schedule," including "proposal-level estimates"; and

(3) "internal company/division software development standards and documented methods and procedures" for "software engineering tools and methods."

Hughes's high risk rating for software capabilities resulted from the agency's concern with respect to Hughes's failure to validate the methods and tools used by the firm to generate its software size estimates and the resulting man power estimates (which were used in the calculation of cost).

The Air Force attempted several times to obtain information concerning the validity and reliability of the methods used in the firm's software planning and estimating. Specifically, in CR (No. 283), the Air Force advised Hughes that:

"To prove the validity of software tools, please take your original [software] LOC [lines of code] estimates (including subcontractor provided LOC) and produce man-hour estimates. Then do the same with your latest LOC estimates. Also, provide all inputs used in your software tools to arrive at the man-hour estimates (i.e., LOC, experience levels, language complexity, test level, etc.)."

Hughes responded that "in generating estimates for new program LOC," the firm used its own "Software Engineering Cost Estimating Guidelines for the Trainer Software Department." According to Hughes, its guidelines, which were included in its original offer, were "the compilation of over 15 years of actual program data and its breakdown by task type." Included in the firm's software guidelines were coefficients--LOC per man-week--for software development in three software languages--High Order, Ada, and Assembly.

The agency considered Hughes's response insufficient and submitted another CR (No. 286) to the firm, as follows:

"Refer to CR No. 283 response: The listed software development coefficients are supported by a claim of 'over 15 years of actual program data.' Please provide actual program data that support these coefficients. For each program include contract number, dates of performance, original estimated LOC, estimated software man-hours, and estimated software development coefficients; and final actual LOC, software man-hours expended, and software development coefficients. Provide substantial information to fully support each of the software development coefficients listed (high order, Ada, and assembly)."

Hughes responded that its claim regarding 15 years of actual program data "is supported by software productivity estimating coefficients, but [deleted]." Hughes explained that "[s]oftware productivity data is more valuable to Hughes than examining original LOC estimates versus final LOC measurements," as the agency had requested, since "productivity data is based [on] [deleted]" and "do[es] not encourage engineers to make programs bigger when they modify them." According to Hughes, "considering only the final LOC tends to mask the cost of changes by hiding it inside software productivity," and "[a]s a result our software estimates [deleted]." Hughes did furnish data for various programs, including a graph "represent[ing] the experienced software productivity in terms of lines of executable source code produced per programmer man-year," with productivity "measured [by LOC per man-hour] at critical points in each program, to revalidate the estimated cost."

The Air Force determined that because of Hughes's failure to furnish the requested information Hughes's software estimates were not validated. Specifically, the agency found that Hughes's software estimating tools were "contractor specific with no validation data, or other information proving the reliability of these tools." The evaluated "lack of validated software planning and estimating tools" resulted in "significant concerns about the reliability of the software estimates." As a consequence, according to the agency, there could be "a larger software effort than what has been scoped by the offeror, which would affect schedule and cost, and may affect the feasibility of the technical approach."

As for Hughes's claim that 15 years of actual program data supported its software development coefficients, the source evaluation team (SET) noted that "when asked to provide this actual program data, the offeror instead provided a [deleted], but "did not provide the data requested." Further, the evaluators considered the information given by Hughes in this regard "vague, incomplete, and contradictory." As an example, the SET noted that:

"The offeror states that their productivity data is based on [deleted]. The offeror then states that their estimates do not [deleted]. Four of the factors we requested were estimated LOC, estimated man-hours, actual final LOC, and actual final man-hours expended for past contracts. This is the data that would be used to compute the estimated productivity and actual productivity for new software development. The offeror failed to provide [deleted]. . . . The [deleted] indicates [deleted] program . . . involving Ada [--the software language required here--] developed under

[Department of Defense (DOD)] restrictions. [Deleted] data points are given for [deleted] program [deleted]. However, there is no information to indicate how much weight should be given to each point [deleted]. Furthermore, there is no way to determine, from the [deleted] points given, what the [deleted] for the entire effort was. Therefore, the offeror's answer does not prove the reliability of their coefficients."

Hughes argues that the historical productivity data the firm provided in its offer and in response to the CRS was sufficient to validate its software estimates. According to the protester, its data properly was based on prior similar programs and was derived [deleted]. Hughes claims that the data requested by the agency such as original estimated LOC versus final actual LOC was [deleted]. According to Hughes, it would have been [deleted] to provide data [deleted], therefore, the protester argues that it should not have been required to [deleted] for evaluation purposes. In addition, Hughes argues that the high risk rating it received for software capabilities was inconsistent with the agency's past performance risk assessment of the firm. Specifically, the protester cites the performance risk assessment group's (PRAG) determination that "[a]ll the data collected on [Hughes] concerning cost indicated no relevant problems existed on past contracts" and that Hughes "is capable of performing with low risk to the government." The protester essentially contends that the PRAG's conclusion in this regard establishes the validity of Hughes's software estimating methodology.

Analysis

In reviewing an agency's technical evaluation, we will not reevaluate the proposal; we will only consider whether the agency's evaluation was reasonable and in accord with the evaluation criteria listed in the solicitation. CORVAC, Inc., B-244766, Nov. 13, 1991, 91-2 CPD ¶ 454. A protester's mere disagreement with the agency's judgment is not sufficient to establish that the agency acted unreasonably. United HealthServ Inc., B-232640; et al., Jan. 18, 1989, 89-1 CPD ¶ 43. Here, based on the record, the agency's high risk rating of Hughes's software capabilities, as it relates to the reliability of the firm's software estimates, appears to be reasonable.

First, we find reasonable the agency's position that the data it requested but which was not furnished by Hughes, such as the original estimated LOC versus final actual LOC, was directly relevant to ascertaining the validity of Hughes's software estimating methodology, and thus the reliability of its software estimates. The agency could

reasonably consider the ultimate accuracy of Hughes's prior initial software estimates as bearing on the validity of the software estimating methodology which produced the estimates. Accordingly, since Hughes's software estimates [deleted], historical data showing the accuracy of the prior initial software estimates was properly viewed by the agency as significant.

Indeed, Hughes's own "Software Engineering Practices Manual" (SEP), submitted with the firm's proposal, recognizes [deleted].

Nor do we believe that the agency was unreasonable in questioning the historical data submitted by Hughes. Our review of the record confirms the agency's position that [deleted] of the [deleted] past contracts cited by Hughes [deleted] clearly similar to the requirement here--a DOD contract using Ada as the programming language--and that the data submitted for [deleted] failed to support the validity of the software estimating coefficient used by Hughes for the contemplated effort here. As discussed by the SET (in the above quote), the data furnished by Hughes for [deleted] indicates the total number of LOC changed, and the productivity rates [deleted]. As also noted by the SET, however, there was no indication of the weight to be given to each of the data points, [deleted]; as a result, the [deleted] for the entire effort could not be determined.² While the protester maintains that the data points on the [deleted] were of "equal (representative) weight" and thus "an [deleted] could reasonably be estimated from those data points," this was not apparent from the submitted data. In any event, the average of the [deleted] given was [deleted] LOC per man-hour, or [deleted] LOC per man-week, which was significantly less than the [deleted] LOC per man-week Hughes used as its software estimating coefficient for Ada here.

Furthermore, Hughes's high proposal risk rating for software capabilities was not inconsistent with its low evaluated performance risk rating. As stated in the RFP, proposal and performance risk were evaluated separately under different criteria (as well as by different evaluators)--proposal risk was to be used to assess the risk associated with an offeror's proposed approach, while performance risk was to

²While the LOC productivity rate the protester used in its software estimates for the effort here, [deleted] LOC per man-week, was approximately equivalent to the [deleted] LOC per man-hour (or [deleted] LOC per man-week) shown for the [deleted] points on the submitted [deleted], there was no indication from the data given that this productivity rate was [deleted].

be used to assess the probability of success of the proposed effort, based on an offeror's present and past performance. The fact that the firm had, in general terms, performed successfully in the past, did not dictate a finding under the proposal risk assessment that the actual software approach proposed here was without risk and that its estimates were reliable.

DISCUSSIONS-SOFTWARE CAPABILITIES

Hughes argues that it was never informed of the insufficiency of its data with respect to software capabilities after its response to CR No. 286 and, further, that if the agency desired data other than that submitted, it was obligated to request it expressly in order for discussions to be meaningful.

The requirement for meaningful discussions with offerors is satisfied by advising them of deficiencies in their proposals and affording them the opportunity to satisfy the government's requirements through the submission of revised proposals. Federal Acquisition Regulation § 15.610(c)(2) and (5); TM Sys., Inc., B-228220, Dec. 10, 1987, 87-2 CPD-¶ 573. The discussions with Hughes concerning the evaluated lack of validation of its software estimating approach were meaningful. The first CR issued in this area requested information "to prove the validity of software tools"; the second CR requested the "actual program data" to support the firm's software estimating coefficients, including original estimated LOC versus final actual LOC, and specifically requested "substantial information" to "fully support" the software development coefficients presented by the firm. These CRs clearly were sufficient to place Hughes on notice of the perceived weakness in its proposal and afford it a reasonable opportunity to satisfy the government's requirements through the submission of a revised proposal. To the extent that Hughes believes that it should have been afforded additional opportunities to revise its proposal after its second BAFO was determined inadequate, there is no requirement that agencies notify offerors of deficiencies remaining in BAFOs or conduct successive rounds of discussions until such deficiencies are corrected. See Honeywell Regelsysteme GmbH, B-237248, Feb. 2, 1990, 90-1 CPD ¶ 149.

COST EVALUATION

Background

The Air Force calculated the GEMPC for each proposal by totaling its estimate of the most probable cost of performing the three categories of required effort-- development, production, and contractor logistics support.

At issue here is the agency's method of estimating the cost of software development, the major task to be performed under the contemplated contract.

The specific starting point for calculating the most probable software development cost was the government's model estimate of the required LOC for the trainer, as broken down into five computer software configuration items --instructor/operator station, student work station, simulator system courseware development system, graphics work station, and simulated test equipment. This estimate, which assumed all new software development and was made prior to the receipt of proposals, was based on the total LOC for the current, similar Airborne Warning and Control System (AWACS) radar trainer, adjusted to reflect the specific, minimum requirements for the JSTARS trainer and the project engineer's prior experience in developing simulation software. The model LOC estimate was then reduced to account for each offeror's proposed use of COTS and reused software--the two areas the agency determined presented the only significant differences among offerors' approaches--to arrive at the net new LOC each offeror had to develop for the system.

The agency then entered the resulting offeror-specific LOC estimates, along with pertinent information in 63 other areas or parameters, into a commercial computer software estimating program, known as the System/Software Estimating and Evaluation of Resources-Software Estimating Model (SEER-SEM), in order to convert the LOC estimates into man-hour estimates so as to arrive at a GEMPC in man-hours for each offeror. Many of these parameters remained constant for all proposals, such as complexity of the programming language (all offerors were required to use Ada),

³While the JSTARS trainer is similar to the AWACS trainer, the agency reports that there are differences. The JSTARS trainer will have more functionality than the AWACS trainer; the JSTARS trainer software will be written in Ada, while the AWACS software was written in the Assembly and Fortran programming languages; and the AWACS trainer software architecture is obsolete.

⁴SEER-SEM is a commercially available software cost, schedule, and risk estimating tool which predicts software costs on the basis of quantitative variables related to such characteristics of the software product as size of the project and functions to be performed. It is widely used within the Air Force, and the agency reports that it has found that SEER-SEM estimates typically fall within 10-15 percent of man-hour estimates calculated using other industry accepted software estimating programs.

requirements volatility (requirements were determined by the Air Force, and changed only as a result of Air Force need), and security requirements. Other parameters varied based on the characteristics of each offeror's proposal and the agency's evaluation of that proposal. Examples of these latter parameters were analyst capabilities and programmers' language experience.

Hughes argues that the cost evaluation was fundamentally flawed because the starting point for the GEMPC for each offeror's proposal was the model LOC estimate, which the protester describes as an estimate based on the government's approach to developing the trainer, but which bore no reasonable relationship to the individual offerors' approaches. The protester maintains that the agency should have started with separate LOC estimates for each offeror, tailored to the particular offeror's approach, and then made the appropriate adjustments. Further, according to the protester, the agency's general approach of reducing its baseline LOC estimate by the number of LOC associated with the use of COTS and reused software proposed by an offeror did not reasonably take into account differences between its own and AAI's lower-level software design approach (i.e., below the configuration item level) and differences in the types of COTS software proposed by each. In addition, Hughes argues that, even if the original model LOC estimate was valid, the agency's estimates of the percent of LOC accounted for by the firms' use of COTS software for the instructor/operator station and student work station items were "questionable." Noting that AAI proposed to use the OS/2 operating system in some areas while Hughes proposed to use DOS/Windows, the protester asserts that the agency's determination that AAI was offering more COTS LOC than Hughes for the above items failed to take into account "the much wider availability of COTS software, tools, and products in the DOS/Windows software operating environment . . . compared to the OS/2 operating system environment."

Analysis

We find no basis to question either the agency's use of a software estimating model or the offeror-specific LOC estimates derived from use of the model. Specifically, we find reasonable the agency's general approach of establishing a common basis for comparison of proposals by using a software estimating model having as its baseline the historical LOC totals for the similar AWACS radar maintenance trainer, adjusted to account for different JSTARS requirements. Hughes has not refuted the premises underlying the agency's approach that a common basis for comparison was necessary, that the AWACS maintenance trainer was the trainer most similar to the JSTARS trainer, and that therefore the AWACS data was the available data most likely

to provide useful LOC estimating relationships. See generally, Newport News Shipbuilding and Drydock Co.; et al., B-254969; et al., Feb. 1, 1994, 94-1 CPD ¶ 198.

Hughes also has not shown that the Air Force's approach to making contractor-specific adjustments was unreasonable. The protester has not demonstrated that to the extent there were any significant differences in Hughes's and AAI's software approaches, these differences were not reasonably accounted for by the numerous parameters in the software estimating model. For example, Hughes has neither provided an explanation as to how the alleged differences in lower-level software design approach and in the types of COTS software proposed in any way demonstrate the unreasonableness of the agency's general approach of subtracting from the model (new) LOC totals the COTS and reused software each offeror proposed, nor has it shown that the agency's specific COTS adjustments were unreasonable. In this regard, even if Hughes is correct that there is more COTS software specifically written for the Windows/DOS environment it proposed than for the OS/2 environment proposed by AAI, Hughes has not rebutted the agency's position that AAI's proposed system necessarily will have access to as much or more COTS software than Hughes's because the OS/2 software system proposed by AAI is 100 percent DOS compatible and can run programs written in DOS (thereby giving AAI access to OS/2 plus DOS/Windows), while Hughes's DOS/Windows system will only have access to programs written for DOS/Windows (and not to those written for OS/2). Further, to the extent the protester believes that the agency should have considered factors beyond those accounted for by the numerous parameters of the model, the protester neither identified those factors nor explained their relevance.

In summary, Hughes has not shown that the Air Force's cost model was flawed so as to call into question the agency's determination that Hughes had significantly underestimated the required LOC, and thus the likely cost of its proposal.

The protest is denied.

Robert P. Murphy
General Counsel