AUTOMATING SUPPLIER EVALUATIONS: USAF ACQUISITION TRANSFORMATION

Authors: Michael Knipper, US Air Force, 50 Chennault Circle, Maxwell AFB AL, 36112, (334)953-0895, michael.knipper@maxwell.af.mil. Timothy Reed, U.S. Air Force, SAF/AQC 1060 Air Force Pentagon, Washington DC, 20330-1060, (703)-588-7061, timothy.reed@pentagon.af.mil

ABSTRACT

The development of automated supplier rating systems has gained popularity in the civilian marketplace over the last few years. Based on the increase of electronic media in the business-to-business environment it has become essential to modify traditional supply chain management (SCM) practices in the commercial sector. The US Air Force (USAF) acquisition community is currently studying Purchasing and Supply Chain Management (PSCM) as a means to improve its own supply chain. One of the ultimate goals of PSCM is to improve the supplier selection process. Two suppliers, which sell the same product with the same specifications and terms, rarely perform at the same level [13]. Many commercial activities have implemented rating scales (e.g., gold, silver, bronze) to evaluate supplier performance in a more timely fashion.

PROBLEM

This research focuses on four of PSCM's 13 key tenets. They include: 1) the automation of routine activities, 2) better management of key suppliers, 3) understanding the supply chain, and 4) rationalizing the supply base. It appears that the private sector has devised many new strategies for success. The use of automated supplier evaluation techniques continues to increase in private industry. For the most part, the USAF acquisition community continues to use time-consuming evaluation techniques in assessing a supplier's performance. This research intends to examine why such a gap exists and provides recommendations on how to close it.

RESEARCH OBJECTIVES

There were five primary objectives of the research: 1) comparing supplier evaluation methods of commercial companies with that of the USAF acquisition community, 2) investigating the success of automated evaluations in the commercial sector, 3) determining best practices for evaluating suppliers in the commercial market, 4) determining what ratings can be quantified during an evaluation, and 5) determining if the Federal Acquisition Regulation (FAR) will prohibit USAF acquisition personnel from transforming the supplier evaluation process.

RESEARCH/INVESTIGATIVE QUESTIONS

This research presents an alternative to the current process. The ultimate goal is to explore whether "commercial" supplier evaluation methods can replace the USAF acquisition pre-award and post-award processes. Further, the objective is to determine which commercial evaluation methods are most successful in selecting the most competent and timely suppliers.

METHODOLOGY

This research was conducted as a case study. The case study included content analysis of the interview responses of eight senior managers from commercial companies in the aerospace or air transportation industries, and two senior managers from government program offices. The interview instrument was devised of questions, which pertained to industry and government, industry only, and government only.

The content analysis of the open-ended responses was aided by the statistical software StatPac (© 2003). After interviewing all ten personnel, the responses were transcribed and transferred into StatPac. The statistical information from StatPac provided the framework for the conclusions and recommendations.

The sampling frame began with Fortune 500 companies that participated in DoD acquisition or currently work in the aerospace industry (e.g. aircraft manufacturing, parts suppliers, air transportation). Some of the companies included Federal Express, United Technologies, Delta Airlines, and Lockheed Martin. On the USAF side, two source selection programs at Wright-Patterson Air Force Base, OH were selected. They included the C-130 Avionics Modernization Program (C-130 AMP) and the Air Force Research Laboratory Enterprise Business System (AFRL EBS). The Government and commercial company interviews were conducted by phone or face-to-face with senior managers involved in strategic sourcing decisions.

EFFORT WITHOUT STRATIFICATION

Pre-award and post-award supplier evaluations require extensive amounts of time for USAF acquisition personnel. During the pre-award evaluation period, contracting officers and members of the past performance evaluation team send out surveys and research contractor past performance data to determine which offerors are most capable of meeting the government requirement. This process can take as long as six months. The current processes for pre-award and post-award collection are too cumbersome in today's acquisition environment [3]. More often than not, the performance risk assessment group (PRAG), which is responsible for pre-award evaluations during source selections, does not establish discriminators between offerors. As a result, the PRAG results are often overshadowed by the technical and cost portions of the source selection.

Businesses that deal with the federal government feel that the Federal Acquisition Regulation allows too much latitude to evaluators, resulting in inconsistent evaluation factors among federal agencies [5]. Additionally, during post-award reviews, the Award-Fee Board (AFRB), which often includes the contracting officer, the program manager, and the logistics manager, requires continuous monitoring of the contractor/supplier in order to determine the award-fee earned. This process typically takes 30 to 60 days. As a result, the contractor is not paid until three months after the work is performed. A recent contracting officer survey reported that award-fees are generally assessed at 80% to 95% of the allotted award-fee pool [14]. This percentage is extremely high for the amount of work and effort expended during the 30 to 60 day evaluation.

The Air Force attempted to use a performance rating system at the Air Logistics Center depots in 1996. It was known as the Blue Ribbon Program (BRP) [2]. It was a best value award system, which authorized contracting officers to award contracts based on factors deemed more important than price. It was a certification rating designed to assess performance on two factors: quality and timely delivery. Contractors were designated "blue ribbon certified" and their credentials for producing designated aircraft parts were continuously monitored. Theoretically, if the BRP firm was within 20% of the

lowest bidder, it could be awarded the contract based on past performance. The program failed. The Air Force cancelled the program because it required excessive documentation for contracting officers to comply with the full-and-open competition requirement of the Competition in Contracting Act [2].

INDUSTRY SOLUTION

Industry has struggled with supplier management just as the USAF has. During the production of the Boeing 747 and 737 airplanes, Boeing supplier failures cost over $1B [11]. To solve problems like these, many commercial firms have devised a paperless media for tracking their suppliers. Many commercial firms use software driven performance ratings that combine qualitative and quantitative measures. One example of the software driven system is Open Ratings (OR). OR gathers opinion data, transaction data, and third-party financial data information from Dun & Bradstreet in computing the overall score (OpenRatings.com, 2002). More specifically, the OR system measures overall performance based on reliability, cost, order accuracy, delivery-timeliness, quality, business relations, personnel, customer support, and responsiveness.

BUYER/SUPPLIER RELATIONS

Supplier selection is the primary task for purchasing managers [10]. This task is becoming more complex in the global business environment today because selection decisions include total cost of ownership (TOC) considerations [6]. TOC includes every aspect associated with the life-cycle of the product. How much will it cost? How much will it cost to maintain? Is the system/product reliable?

The traditional criteria of price, quality, delivery, and speed are changing to include financial data. Due to proprietary information and traditional thinking, exchanging information was extremely uncommon between firms. Figure 1 summarizes the changes in buyer/supplier relations [7].

Figure 1

Factor	Traditional Approach	Strategic Approach
Desired relationship w/ supplier	Short run Multiple sourcing Antagonistic	Long run Single-sourcing Cooperative
Objective	Tactical Objective	Strategic Objective
Selection criteria	Price Specified quality Delivery speed	Total cost Total quality management Service, Financial stability, present and future technological capabilities, organizational culture and strategy, environmental concerns, international supply, supplier record, supplier customer portfolio
Locus of the Purchasing decision	Different departments	Multifunctional teams w/ purchasing, engineering, marketing, and quality
Supplier Evaluation	Unstructured according to minimum specified values	Structured Evaluation points/weights in total cost models, objective

INTERVIEW DATA ANALYSIS & RECOMMENDATIONS

Primary Research Question 1 asked "can USAF acquisition personnel use supplier ratings in lieu of conducting award-fee boards or past performance evaluations?

Interview Question (IQ) 10 asked "How is post-award evaluation maintained by your firm/program office?" Many of the companies interviewed reported having instituted scorecard rating systems to evaluate their suppliers after contract award and others are in the early stages of developing scorecards. In contrast, the government uses previous source selections and headquarters command or air staff guides to model their evaluations.

IQ 11 asked "is the process automated or manual?" What is the role of purchasing personnel in this process? The responses for this question were somewhat mixed. Many companies used manual retrieval processes while others had databases which could be queried automatically to obtain quality and delivery information. In general, the role of purchasing personnel was limited to data retrieval.

IQ 4C (C=Contractor Only) asked "do you maintain one rating for each supplier which covers all contracts or are they given a rating which is based on their performance on one particular contract?" Many companies are still trying to determine if it is necessary to give ratings for each contract or if an across the board approach is sufficient. Honeywell, Delta, and Boeing use across the board roll-ups for divisions. Raytheon, FedEx, UTC, and Lockheed Martin can retrieve information on particular contracts but often roll-up their information to the division level.

IQ 16 asked "what 3 things could you have done better in pre-award and post-award evaluations knowing what you know now?" This question resulted in a variety of answers. The four most prevalent were: 1) up-front work, 2) objective criteria or a standard process, 3) firm specifications, and 4) eliminating ethical concerns.

Our recommendation is that the USAF acquisition community devise an automated scorecard system to track suppliers on contracts above $25K. If such a system were intended for small purchases it could become extremely time consuming on small repetitive buys. Although my research was limited to two program offices at the product center level, an automated system could also provide information for depot and operational buying units. Most commercial companies are moving in the automated scorecard direction and those that have implemented scorecard ratings perceive them to be extremely effective. Potential suppliers as well as current suppliers are welcoming the feedback they receive from these companies as a way to see how well they're doing.

The implications for the USAF are that a rating system would have to be incorporated with its standard procurement system. The system would need visibility of all individual contract ratings for a particular supplier and an aggregate rating for pre-award assessments. If there are post-award implications, then only the ratings for a specific contract would be queried. The caveat with any government system would be to maintain subjective ratings. Quantitative ratings are effective but often result in a different outcome than desired. The government interviewees seemed to be hesitant in making all ratings quantitative. In one interview, the program manager remembers a General saying, "if we only use quantifiable numbers, that's all the contractor will strive for on paper" [4]. Research supports this fear: "Whether dealing with monkeys, rats, or human beings, it is hardly controversial to state that most organisms seek information concerning what activities are rewarded, and then seek to do (or at least pretend to do) those things, often to the virtual exclusion of activities not rewarded" [9].

Subsidiary Question 2: What are the similarities between USAF rating categories and commercial rating categories?

IQ 1 asked "what are the categories/factors of assessment used for evaluating past performance? How were these categories constructed (i.e. industry standard, from within)? The most common factor was on-time delivery. Quality and past performance were also important. In this case some companies mentioned price or cost, and although they are not truly the same, the terms were joined for simplification. How the categories were constructed was not significant.

IQ 2 asked "is your rating system quantitative (e.g. late delivery in terms of # of days) or qualitative (e.g. quality performance)? This question produced an interesting difference between commercial companies and government organizations. The commercial companies expressed a desire to make their rating categories as quantitative as possible. However, the commercial companies explained that a mixture of the two was more common. On the contrary, the government organizations explained the importance of qualitative (subjective) ratings.

IQ 3 asked "how many days on average does it take you to award a contract (e.g. clock starts when a requirement or funding document is received)?" All of those interviewed asked for a dollar value in order to answer the question. A theoretical $50M cradle-to-grave buy was used for interview purposes. Answers ranged from 3 weeks to one year. Overall, the commercial companies were much faster.

IQ 4 asked "has your pre-award evaluation process changed over the last 5 years?" Most companies stated the process had become more automated. Delta Airlines stated that their process has changed in the way they collect and distribute information to a supplier. Companies such as United, Honeywell, and FedEx are using reverse auctions with pre-qualified vendors. Raytheon stated that they have changed from a tool and process standpoint…"we've gotten more electronic and web-enabled" [12]. The government personnel stated that changes over the last five years have been minimal.

IQ 7 asked "would you categorize most of your contracts as fixed-price, incentive based, or cost-plus?" This question indicated that most commercial firms use fixed-price contracts. The government offices reported cost-type arrangements as the most common. Most of it depends on what type of buying an organization is engaged in.

IQ 8 asked "is the importance of price tailored for each acquisition?" The most common response was yes, but it depends. In most cases there was an increased emphasis on non-price related factors.

IQ 9 asked "is price the most important factor in evaluation?" Half of those interviewed answered "no" to this question. However, those that said it was a best value approach were grouped with the "no" answers. All ten interviewees said that price was a factor in every buy, but not necessarily the most important factor.

IQ 17 asked "how much training is given in-house regarding supplier evaluations?" A few of the firms had some type of training to discuss supplier evaluations; however, they consolidated it with the annual training for purchasing and sourcing personnel.

The USAF rating categories are in line with that of commercial companies. There is little evidence to indicate that there is a distinct difference between the commercial company and USAF evaluation categories. The only difference is that the USAF system accounts for additional qualitative ratings.

On-time delivery and quality were the most common factors examined during an evaluation. Where the government falls short is in its ability to automatically track the delivery or receipt of an item. At some of the commercial companies the receiving dock automatically inputs delivery information into a database. This information can be queried to assess the delivery performance of a supplier. The same can be said about quality at inspection. The receiving/inspection personnel will input quality defects or acceptance rates. In the government environment purchasing personnel are unaware of delivery until they receive an invoice or the user calls to complain about the quality of a commodity or service.

In terms of award-fee boards it was extremely hard to compare the categories between commercial companies and USAF program offices. Commercial companies predominately use fixed-priced arrangements. Very few of them have incentives to exceed the minimum requirement of a contract other than more business. The one recommendation for improvement in the post-award government categories would be to become more quantitative in the three areas of *cost performance*, *time of delivery*, and *schedule.*

Subsidiary Question 3: Can cost-avoidance savings be quantified if either were replaced?

IQ 3 asked "how much time (hours, days) is required to complete a pre-award evaluation?" For a $50M project I got answers ranging from 3 weeks to one year. Overall, this answer was dependent upon the urgency of the requirement.

IQ 5 asked "have you been able to quantify any savings over the last 5 years with your evaluation process?" Many of the answers for this question were not put in terms of dollars. UTC reported a $1B savings through 2001 on the product side to Wall Street and a $700M level savings by the year 2004.

IQ 12a asked "how much time (hours, days) is required to complete a post-award evaluation?" This question was extremely dependent on how automated a company's evaluation system was. Delta Airline's process took about 30 minutes for each of its 50 strategic suppliers.

IQ 12b asked "how often are post award evaluations performed (never, monthly, quarterly)?" Some of the companies can update this information at the first of the month but the typical requirement was to update supplier rating information quarterly. At Boeing the process takes about 30 minutes each month to compile the data into a 12 month moving average.

IQ 19 asked "how many individuals typically participate in a pre-award or post-award evaluation for $50M?" "20 or 30 people wouldn't surprise me, but I would definitely say that we use a core of 6 to 10 people" [12]. "We would use a core of four people: a commodity lead, a quality representative, and an engineering representative; however, it would definitely touch a lot more" [8]. The government teams were both estimated at around 15 to 20 people. These answers did not produce significant inconsistencies between the government and commercial companies.

Government cost-avoidance might be quantified if both source selections and award-fee evaluations were less time consuming. When source selection questionnaires concerning supplier past performance are sent to government contracting officers and program managers they are returned with quality, on-time delivery, and management information. This effort is time consuming for a PRAG or source selection team. This was evidenced by the $15M AFRL EBS acquisition, which took 1 year to complete. If quantitative and qualitative information could be pulled from a supplier scorecard database then questionnaires would become unnecessary. Additionally, in the case of the C-17 award-fee board over 100 man-hours were spent each quarter in assessing McDonnell Douglas's (now Boeing) performance [4]. Some of this time would be saved if ratings were automated.

Subsidiary Question 4: What acquisition reform initiatives or regulations would aid or hind USAF application of a revised performance evaluation system?

IQ 1G (G=Government Only) asked "what is your perception of the FAR's restrictions on the pre-award evaluation process?" The government personnel that were interviewed did not find the FAR restrictive when applied to source selection evaluations. "Most of the guidance that we used in carrying out our PRAG was driven down from Air Force Materiel Command Headquarters or air staff" [4]. "Most of the procedures aren't in the FAR but come down from headquarters" [1].

IQ 6C. (C=Contractor Only) asked "what regulations does your company have which outline the pre-award evaluation process?" Procedures and policies are not lengthy documents in comparison with the Federal Acquisition Regulation. "Because it's web-based there isn't a ton of procedures around it. It's an internal document that is published or sent via e-mail. Buyers do have a desktop guide that is 10 pages long" [12].

My research fails to provide evidence that reform initiatives or FAR regulations would impede a more automated performance evaluation system. In fact, the FAR is not as restrictive as was hypothesized. Most of the guidance for evaluations comes from Air Force Materiel Command Contracting Division (AFMC/PK) or the Deputy Assistant Secretary (Contracting) of the Air Force (SAF/AQC). Much of the guidance is "lessons learned" from previous source selections.

RECOMMENDATIONS FOR FUTURE RESEARCH

Future research should include a comparison of the actual scorecards that companies have developed. During this study, we received two example scorecards, one from FedEx and one from UTC. If the USAF is going to implement scorecards it should obtain feedback on the most and least desirable rating categories. Another recommendation would be to sample the suppliers that have scorecards on file with these companies. Although the companies felt that the suppliers welcome this type of rating system, it would be interesting to know if this holds true. Additionally, supplier qualification and certification plays a huge factor in establishing long-term relationships. The USAF is currently investigating ways to recognize superior performers without it being perceived as favoritism for future acquisitions. Lockheed Martin's STAR supplier program seems to be an effective way of building pride in workmanship for its suppliers. Further research could focus on the "best practices" of various supplier award programs.

ADDITIONAL FINDINGS FROM INTERVIEWS

Many commercial firms have established sourcing departments within their company. These departments are responsible for all major source selections. One proposal is for source selection offices at the product centers to function as more than advisors. This organization could become a "sourcing team". Theoretically, the same individuals that work the past performance portion for the F-16 avionics upgrade would work the past performance portion of the B-2 engine upgrade.

Many suppliers in the automotive industry are held to the same process/quality standards as aerospace firms. Honeywell mentioned that many of these suppliers are no longer doing business with Ford or GM because of optimization (i.e. supplier reductions). These suppliers have many of the capabilities and resources that could be applied towards DoD efforts. It would be interesting to see if these companies could help alleviate some of the problems associated with the DMS (Diminishing Manufacturing Sources) challenge that the DoD currently faces.

REFERENCES

[1] Air Force Research Laboratory (AFRL) Enterprise Business System, Aeronautical Systems Center, Wright-Patterson AFB, OH. Performance Review and Assessment Group Chairman, Personal Interview. 12 Jan 2003.

[2] Ambrose, Matthew H. *Past Performance in Supplier Certification Programs: A Study of Current Certification and Incentive Practices in Certified Supplier Programs*. Naval Postgraduate School, Monterey CA, June 1997 (AD-A333461).

[3] Beausoleil, J. "Past Performance is Not Working," *Contract Management*, 40: 28-31 (2000).

[4] C-130 Avionics Modernization Program, Aeronautical Systems Center, Wright-Patterson AFB, OH. Performance Review and Assessment Group Chairman, Personal Interview. 12 Jan 2003.

[5] Clipsham, N. "Appraising performance appraisal systems," The Military Engineer, 591: 31-32 (1998).

[6]] Ellram, Lisa M., "Total Cost of Ownership: An Analysis Approach for Purchasing," *International Journal of Physical Distribution & Logistics*, 25(8): 4-23 (1995).

[7] Goffin, K., Szwejczewski, M., and New, C. "Managing Suppliers: When Fewer Can Mean More," *International Journal of Physical Distribution & Logistics*, 27: 422-436 (Issue 7/8 1997).

[8] Honeywell Corporation, Phoenix, AZ. Director of Supplier Quality and Supplier Health Management, Telephone Interview. 22 Nov 2002.

[9] Kerr, Steven. "On The Folly of Rewarding A, While Hoping for B," Academy of Management Journal, 18(4): 769-783 (December 1975).

[10] Monczka, Robert M., Robert B. Handfield, and Robert J. Trent. *Purchasing and Supply Chain Management* (2nd Edition). Cincinnati, OH: South-Western, 2002.

[11] Seungwook Park, Darryl D. Wilson, and Janet L. Hartley. "Quality Management Practices and Their Relationship to Buyer's Supplier Ratings: A Study in the Korean Automotive Industry," *Journal of Operations Management*, Vol. 19, No. 6: 695-712 (2001).

[12] Raytheon Corporation, Lexington, MA. Vice President of Enterprise Supply Chain Management, Telephone Interview. 6 Dec 2002.

[13] Smith, Stanford A. "Third-Party Supplier Ratings," *Purchasing,* 129: 40-41 (22 December 2000).

[14] Snyder, Thomas J. "Analysis of Air Force Contract Implementation," *Air Force Journal of Logistics*, 25: 14-21 (Summer 2001).