



ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Exploring the Efficacy of the Government's Current Use of Past Performance Information

3 December 2013

Dr. Timothy G. Hawkins, Assistant Professor

Graduate School of Business & Public Policy

Naval Postgraduate School

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.



The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Abstract

Since its inception via the Federal Acquisition Streamlining Act of 1994, contractor past performance is intended to be an important evaluation criterion in federal source selections. In order to reduce performance uncertainty, procurement officials must record contractor performance evaluations in a central database. However, reports of ubiquitous problems raise questions of the integrity of ratings and the utility of the evaluations. From a literature review, several factors affecting the efficacy of past performance evaluations are identified. These factors are combined in a comprehensive conceptual model explaining past performance efficacy. Exploratory, qualitative data preliminarily confirms the hypotheses. Key antecedents include the following: rating justification quality; contractor surveillance; multi-rater dissonance; perceived accuracy; evaluator role overload; fear of supplier dispute; perceived fairness; sufficiency of requirement definition; evaluator turnover; relationship quality; and buyer–supplier communication frequency, bi-directionality, and formality. From these findings, important managerial and theoretical implications are drawn and future research directions are identified.

Keywords: Contractor, Past Performance, Supplier Performance Evaluation, Contractor Performance Assessment Report



THIS PAGE INTENTIONALLY LEFT BLANK



About the Author

Timothy G. Hawkins, Lt Col, USAF, PhD, is an assistant professor in the acquisition area at the Naval Postgraduate School in Monterey, CA. He directs and teaches in the Strategic Purchasing program aimed at transforming military procurement. He has 19 years of sourcing experience in industry and government. Lt Col Hawkins has published articles on opportunism in buyer–supplier relationships, performance-based logistics, collaborative pricing, and electronic reverse auctions in scholarly publications such as *Journal of Supply Chain Management*, *Journal of Business Research*, *International Journal of Logistics Management*, *Industrial Marketing Management*, *Journal of Business Ethics*, *Supply Chain Management: An International Journal*, *Journal of Marketing Channels*, *Air Force Journal of Logistics*, and *Journal of Public Procurement*. His current research interests include electronic reverse auctions, procurement ethics, buyer–supplier relationships, strategic sourcing, services procurement, and supplier performance management.

Timothy G. Hawkins, Lt Col, USAF, PhD
Graduate School of Business & Public Policy
Naval Postgraduate School
Monterey, CA 93943-5000
Tel: (831) 656-7647
Fax: (831) 656-3407
E-mail: tghawkin@nps.edu



THIS PAGE INTENTIONALLY LEFT BLANK





ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Exploring the Efficacy of the Government's Current Use of Past Performance Information

3 December 2013

Dr. Timothy G. Hawkins, Assistant Professor

Graduate School of Business & Public Policy

Naval Postgraduate School

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the federal government.



THIS PAGE INTENTIONALLY LEFT BLANK



Table of Contents

| | |
|---------------------------------------|----|
| Introduction | 1 |
| Literature Review | 5 |
| Past Performance..... | 5 |
| Supplier Performance Evaluation | 6 |
| Agency Theory | 8 |
| Organizational Behavior | 10 |
| Channel Communication | 13 |
| Social Exchange Theory | 14 |
| Power/Dependence..... | 15 |
| Methodology | 17 |
| Qualitative Data Analysis | 17 |
| Data Collection | 17 |
| Data Analysis | 17 |
| Sample | 17 |
| Results..... | 18 |
| Discussion | 28 |
| Managerial Implications..... | 29 |
| Theoretical Implications..... | 32 |
| Study Limitations | 32 |
| Future Research Directions..... | 33 |
| Conclusion | 33 |
| References | 35 |



THIS PAGE INTENTIONALLY LEFT BLANK



List of Figures

| | | |
|-----------|------------------------|----|
| Figure 1. | Conceptual Model | 16 |
|-----------|------------------------|----|



THIS PAGE INTENTIONALLY LEFT BLANK



List of Tables

Table 1. Informant Demographics 18



THIS PAGE INTENTIONALLY LEFT BLANK



Exploring the Efficacy of the Government's Current Use of Past Performance Information

Introduction

In November 2012, the Air Force cancelled a \$1 billion acquisition program to purchase and integrate a new logistics enterprise resource planning (ERP) system. This system was to replace 240 legacy systems and improve logistics costs and performance. Despite its positive evaluation of the systems integrator (Computer Sciences Corporation [CSC]) and of the commercial software provider (Oracle) at contract award, the Air Force attributed the program's failure to CSC's lack of capability and skills to perform (Reilly, 2012). Could this risk have been discerned during proposal evaluations—perhaps informed by better past performance information? While the impact of the past performance evaluation is speculative, the question, more ubiquitously applied to all of our critical source selections, holds merit.

Industrial buyers labor to avoid the deleterious effects of the laws of agency. In industrial buying, the supplier serves as an agent to the principal (buying organization). Substantial effort is dedicated to avoid adverse selection and moral hazard. *Adverse selection* encompasses the risk of selecting an incapable supplier that otherwise misrepresents itself as capable, while *moral hazard* is the vulnerability to acts of supplier opportunism (Eisenhardt, 1989)—behavior that is self-interest seeking with guile (Williamson, 1975). For example, supplier opportunism could include shirking quality, obfuscating the truth, withholding information, lying, cheating, and breaching contract terms (Wathne and Heide, 2000).

In their buying efforts, government agencies incur significant transaction costs attempting to write all-inclusive contracts and to monitor contractor performance in order to thwart supplier opportunism. These costs of contracting are substantial given the magnitude of contracted goods and services. In fiscal year (FY) 2010, the federal government awarded more than 5.9 million contract actions worth over \$538 billion (Federal Procurement Data System–Next Generation [FPDS–NG, n.d.]). More transaction costs are incurred attempting to mitigate information asymmetries, thereby avoiding adverse selection, by requiring that past performance be an evaluation criterion for contract award. The logic is that by better informing source selection decisions, better best value selections will occur. Integrally related is the contractor's performance; if



performance levels are assessed and recorded, and if this information is available to contracting officers during a future source selection, conventional wisdom holds that contractors will work harder to ensure satisfactory (or better) performance.

In U.S. federal government contracting, agencies are required to consider past performance information as an evaluation factor in formal source selections. Necessarily, then, agencies must collect and report contractor past performance information from certain government contracts. However, there are many concerns that the past performance evaluations/ratings are not properly, timely, or accurately completed. From 2007 to 2010, overdue assessments grew from 5.3% to 10.1% of total assessments required (Contractor Performance Assessment Reporting System [CPARS] Metrics, n.d.). In FY10, 327,111 reports were due, and 257,908 were completed. In 2009, the Government Accountability Office (GAO) estimated that only 31% of contract actions requiring CPARS reporting had completed reports. Reports often lack sufficient information to support ratings (e.g., how the contractor met, exceeded, or failed to meet requirements) necessary to withstand a legal challenge, or do not include a rating for all performance areas (Office of Federal Procurement Policy [OFPP], 2011). Additionally, throughout the rating process, raters often inflate ratings in order to avoid conflict with the contractor (GAO, 2009). The FY10 CPARS data shows quite a low proportion (1.5%) of reports indicating that the rater would not award another contract to the contractor (CPARS Metrics, n.d.).

A related matter pertains to ongoing contractor performance management during the performance of a contract. Due to the impressive effects on buyer performance (Cormican & Cunningham, 2007), supplier performance management (SPM) is an essential best practice in business-to-business sourcing (Gordon, 2008; Talluri & Sarkis, 2002). Despite the demonstrated value of SPM systems in the for-profit sector, the government lacks a coherent strategy and a consistent means to manage contractor performance. A recent study compares the usage rate of SPM systems among best-in-class firms from the for-profit sector (53%) to the public sector (all levels of government—32%; Dwyer, 2011). Whereas contractor performance is closely measured and managed for weapon system acquisitions, the management of contractor performance on service contracts—where the Department of Defense spends most of its funds—is often deficient and inconsistent (GAO, 2001). The government is not alone; the for-profit sector also fails to properly resource service contracts, and consequently, fails to manage service suppliers well (Ellram, Tate, & Billington, 2007).



It is no wonder that government buyers have achieved nearly double the savings (28%) compared to their for-profit sector counterparts on sourcing improvement projects (Husted & Reinecke, 2009). Clearly, there is a plethora of opportunity for the government to improve. In fact, when compared to for-profit sector sourcing, the government lags in all ten dimensions of sourcing (Husted & Reinecke, 2009). The government's void of SPM might explain the variance in raters' ability to efficiently conjure sufficient facts to support a past performance assessment/rating. The obvious question then becomes, why does the government restrict the purpose of past performance information solely to informing future source selections? Is it worthwhile to integrate past performance with a system to manage contractor performance during the contract (versus after contract performance, or once per year)?

Admittedly, the government, in delivering services to the public, serves more masters than does its commercial counterparts. While the government is responsible for obtaining fair and reasonable prices, sometimes this is compromised by other duties such as conducting business transparently, ensuring fairness to contractors, and supporting socio-economic goals (Federal Acquisition Regulation [FAR] 1.102-2). On the one hand, the government needs a means to (1) hold sub-optimally performing contractors accountable, (2) protect itself from adverse selection (i.e., reduce procurement risk), and (3) motivate high supplier performance in order to reap the most value from contracted efforts. On the other hand, contractors are entitled to a fair performance assessment (Graham, 2011). The challenge is figuring out how to provide a fair, accurate assessment in an efficient manner.

Unreliable or inaccurate past performance assessments can harm contractors' reputations and can bias source selections resulting in adverse selection. If past performance information is not reliable, and if contracting officers and evaluators do not (or cannot) use the information to discriminate between competitive proposals (Kelman, 2010), the effort of collecting and reporting the past performance information is squandered. Likewise, the effort of evaluating and documenting inaccurate past performance information during source selections is wasted. Federal contract managers are already overworked (GAO, 2009) and understaffed (GAO, 2001); therefore, continuing to consume time on a fruitless task would be futile.

While the GAO (2009) suggested that assessments and ratings are inflated, the degree of inflation is unknown. Evidence suggests that the magnitude of distortion is high—so much that contracting officers, evaluators, and source selection authorities rarely use past performance information as a meaningful discriminator between proposals. In order to determine whether this



seemingly vacated faith is warranted, the degree of distortion needs to be assessed. The extent of distortion will tell us whether the reporting system and policy needs to be abandoned, adjusted, or left intact. Notably, we don't know how much effort (man-hours) by all parties involved is consumed in completing a past performance evaluation. If the effort is significant, and the resultant information is of little value, policy-makers should revisit the policy and its implementing systems. Additionally, given the current environment of extreme budget reductions, the prospect of achieving savings through improved contractor performance via SPM should be explored.

The purpose of the research, therefore, is to explore the efficacy of the government's current use of past performance information. The intent is to diagnose alleged weaknesses and to explore potential improvements. The following research questions are addressed:

1. Are past performance reports useful? How so, or why not?
2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?
3. Why do reviewing officials change the ratings of the evaluator (assessing official)?
4. How many man-hours does a completed past performance evaluation/rating, on average, consume?
5. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?
6. Why do past performance evaluations/ratings lack sufficient justification/supporting information?
7. Why are past performance evaluations sometimes inaccurate?
8. Do contracting officers use past performance evaluations/ratings to manage contractor performance throughout the contract, or is reporting past performance merely done to comply with the FAR (i.e., effectiveness versus compliance)?

The answers to these eight questions should help diagnose the efficacy of the government's current collection and use of past performance information. The remainder of this paper is organized in the following manner. First, a literature review is presented describing the emergence of our conceptual framework and proposed hypotheses. Next, the study presents the research design and methodology. Lastly, discussion, limitations, implications, future research directions, and conclusions are offered.



Literature Review

Similar to the findings of Ashworth, Boyne, and Walker (2002), we were unable to find a single comprehensive theoretical framework explaining the efficacy of collecting and using supplier past performance information. Such a complex phenomenon can only be explained by synthesizing multiple theories such as those found in the management, marketing channels, supply chain management, and organizational behavior domains. Specific, relevant theories include agency theory, organizational behavior, channel communication, and social exchange theory. Before discussing each theory, the foundation is set by discussing (1) the government's past performance policies and (2) a parallel scheme used in the for-profit sector—SPE.

Past Performance

In U.S. federal government contracting, agencies are required to consider past performance information as an evaluation factor in source selections exceeding the simplified acquisition threshold, \$150,000 (FAR Part 15)—unless the contracting officer documents a reason not to do so. Necessarily, then, agencies must collect and report contractor past performance information from government contracts (FAR Part 42) surpassing certain dollar values (weapon systems, \$5 million; operations support, \$5 million; services, \$1 million; information technology, \$1 million; healthcare, \$100,000; fuels, \$100,000; construction, \$650,000; and architect-engineering services, \$30,000). The FAR defines past performance information as

relevant information, for future source selection purposes, regarding a contractor's actions under previously-awarded contracts. It includes, for example, the contractor's record of conforming to contract requirements and to standards of good workmanship; the contractor's record of forecasting and controlling costs; the contractor's adherence to contract schedules, including the administrative aspects of performance; the contractor's history of reasonable and cooperative behavior and commitment to customer satisfaction; the contractor's reporting into databases; the contractor's record of integrity and business ethics, and generally, the contractor's business-like concern for the interest of the customer. (FAR Part 42.1501)

It is important to note that in keeping with the government's core goal of transparency and fairness (FAR 1.102), contractors must be afforded the opportunity to comment on the government's assessment of past performance, and any disagreements must be resolved by a reviewing official one level above the contracting officer. Additionally, in an emerging area of law, contractor past performance assessments are increasingly subject to the Contract Disputes Act



of 1978 (Lord, 2005). While the courts will not yet direct a particular rating, they will require agencies to adequately support assessments/ratings with sufficient facts. This written justification consumes significant time from the raters, contractors (i.e., rebuttals), and approving officials—as does adjudicating a claim should an assessment/rating be disputed. As further incentive to conceal true performance, program officials will go to extraordinary lengths to protect their programs. A poorly performing contractor can signal a troubled program, increasing the threat of cancelation (GAO, 2009). Other reasons that truthful performance is not reported include a desire to maintain relations with the contractor, difficulty attributing performance problems to the contractor or the government, deficient oversight of contractors, deficient contract administration, and the government’s lack of contractor performance management (GAO, 2009).

Supplier Performance Evaluation

Supplier performance management (SPM) systems became popular in the 1950s (Wieters & Ostrom, 1979), and now SPM is an essential best practice in business-to-business sourcing (Gordon, 2008; Talluri & Sarkis, 2002). SPM is “the process of evaluating, measuring, and monitoring supplier performance and suppliers’ business processes and practices for the purposes of reducing costs, mitigating risk, and driving continuous improvement” (Gordon, 2008, p. 4). SPM systems are used to (1) prioritize supplier improvement activities, (2) focus management attention on critical suppliers, (3) support supplier selection decisions, (4) communicate dissatisfaction with supplier performance, (5) communicate performance expectations to suppliers, (6) document historical performance, (7) inform the purchasing department of supply base performance, (8) influence suppliers, and (9) continuously improve (Schmitz & Platts, 2003). “Performance based systems maximize the use of data, which is then used to convey specific improvement targets, set goals, monitor performance, and evaluate that performance” (Giunipero & Brewer, 1993, p. 39).

It is not surprising that buying firms closely measure their suppliers’ performance when 50%–70% of their revenue is spent on goods and services to support the sales (Monczka, Handfield, Giunipero, & Patterson, 2011). Measuring supplier quality is critical since the cost of poor quality ranges from 10% to 25% of sales, and the cost of poor *supplier* quality ranges from 25% to 70% of the cost of poor quality (Gordon, 2008). Commercial SPM systems—often web-based and at least partially automated—encompass means to measure, rate, and rank suppliers. In 2002, more than half (54%) of for-profit sector buyers did this continuously (Simpson, Siguaw, & White, 2002), and two-thirds of buyers ranked their suppliers based on performance. A more recent study reported a drastic increase in supplier performance measurement and



ranking, showing that 97% of firms use a periodic supplier scorecard or assessment for direct materials (CAPS Research, 2011).

SPM pays off; a study by the Aberdeen Group (2005) found that supplier performance of companies with an SPM system improved significantly more than did the supplier performance of firms with no SPM system. Specifically, firms using an SPM system realized 10% greater price savings, 12% better on-time delivery improvement, four times greater quality improvement, and a 4% greater improvement in service. One large telecommunications firm realized a 290% reduction in the number of suppliers and a 260% reduction in the value of inventory held due to an SPM system (Cormican & Cunningham, 2007). Another study (Limberakis, 2011) found that “best-in-class” buyers (1) are much more likely to benchmark supplier performance against others in the same industry, (2) achieved substantially higher percent on-time delivery (88% versus 48% for “laggards”), and (3) transacted with suppliers that experienced fewer catastrophic failure (2% versus 5% for other buyers). Of the best-in-class buyers, 63% had a supplier benchmarking and performance monitoring information technology system in place. Additionally, the use of a performance evaluation program increases the strength of the relationship between suppliers’ process innovativeness and the buyer’s performance benefits (Azadegan, 2011). The use of an SPM system was also found to improve buyer–supplier relationships (Prahinski & Benton, 2004). Prahinski and Fan (2007) found that the frequency and content of feedback increase the suppliers’ commitment to the buyer, which, in turn, increases supplier performance. Denali Consulting group found that SPM can yield a 3% to 6% cost reduction in total supply chain costs via continuous improvements (Minahan, 2007). A study by CAPS (Monczka, Choi, Kim, & McDowell, 2011) of eight firms found that supplier performance measurement is one of five critical components of effective supplier relationship management (SRM), and that SRM enables vast positive results such as the following: overhead cost reductions, process improvements, increased visibility into actual costs (versus price), year-over-year cost reductions, millions of dollars in savings, product launches on time and on cost, shorter new product development times, total cost reductions of 12%, and quality improvements. As such, all leading purchasing textbooks devote a section to SPM (Benton, 2010; Burt, Dobler, & Starling, 2003; Leenders, Johnson, Flynn, & Fearon, 2006; Monczka, Handfield, Giunipero, & Patterson, 2011; Rudzki, Smock, Katzorke, & Stewart, 2006; Trent, 2007). Not surprisingly, SPM is a core competence of chief procurement officers (Kern, Moser, Sundaresan, & Hartmann, 2011).

Most SPM processes used by buyers integrate subjective and objective evaluations (Simpson et al., 2002; Hald & Ellegaard, 2011). It is assumed that these assessments are accurate; however, as Gordon (2008) pointed out, even



the seemingly most-objective performance parameters, such as percent on-time delivery, can be subjective. The supplier evaluation process has rarely been examined, and social and organizational biases have been ignored (Purdy & Safayeni, 2000). Hald and Ellegaard (2011) found that supplier evaluations are shaped and reshaped throughout the evaluation process. They discovered performance data instability as captured in ERP databases. They also found that evaluations were derived by condensing a larger set of performance information to a smaller, more manageable set of numbers. Buyers also commonly use multiple evaluators to rate supplier performance (Buffa & Ross, 2011; Hald & Ellegaard, 2011). Buffa and Ross (2011) noted the importance of supplier evaluation by functionally heterogeneous evaluation teams. Subjective measures among multiple raters invite dissonance in ratings and opinions—either on the same performance observations or across different instances of performance (Buffa & Ross, 2011). Similarly, Perkins (1993) noted that the different members of the buying organization’s procurement team perceive the supplier’s value delivery differently. While Buffa and Ross (2011) offered an ex post means to accommodate variance among multiple evaluators, there remains little explanation as to systemic sources of the variance. Hence, are there factors that can be managed to mitigate performance evaluators’ dissonance? Additionally, the degree of internal dissonance of supplier evaluations has not yet been examined. Hald and Ellegaard (2011) also reported that performance ratings are sometimes negotiated with suppliers when the accuracy is challenged. However, no one has explored why buyers decide to change their evaluations. Additionally, evaluations are only as good as the data recorded by surveillance; yet, instances of surveillance may not reveal true performance levels (Purdy & Safayeni, 2000).

Agency Theory

This research acknowledges multiple perspectives of agency theory as it applies to industrial exchange. The first perspective views the hired supplier as an agent to the buyer to achieve the buyer’s objectives. The second perspective examines the buyer internally acknowledging that the buyer is comprised of multiple agents to itself. For instance, employees working in procurement, logistics, financial management, engineering, end users of suppliers’ goods and services, and program management represent distinct interests within the firm. Agency theory wrestles with two problems: (1) conflicting interests between principal and agent and (2) difficulty and cost associated with monitoring agents, and the associated uncertainty for not having perfect information (Eisenhardt, 1989).



Beginning with the second perspective, using multiple raters within an organization to evaluate supplier performance can create conflicts of agency. In the case of past performance evaluations, evaluators of performance serve as agents to multiple principals—their employing government agency, their local organization or unit, and the taxpaying constituents. Problems of agency arise when agents' self-interests differ from his or her employer's goals (Bergen, Dutta, & Walker, 1992). Two theories of not-for-profit organizations support self-interested pursuits of agents. Budget-maximization theory (Niskanen, 1968) follows the utility maximization model of rational human behavior to posit that bureaucrats unable to seek greater compensation will instead be motivated to increase their budgets in order to increase their power. In contrast, the bureau-shaping model relies less on the assumption of utility maximization to posit that public managers develop a sense of ownership of their agencies and shape them to satisfy personal utilities (Barberis, 1998). Rather than simply enlarging the organization or accumulating power, bureau-shaping predicts other managerially desired outcomes such as reducing personal risk and increasing access to centers of power in ways that do not unduly increase the scope of the problems under their responsibility. Both models agree that self-interest motivates public managers to accumulate power for personal gain. These self-interests can conflict with that of employers, thus, creating problems of agency. For example, often, evaluators fail to properly monitor a contractor's performance. If the contractor's performance did not meet requirements, rather than rate the contractor as unsatisfactory, the evaluator might inflate the rating to avoid a dispute—conflict that would unveil the evaluator's negligence. Agency theory holds that once the principal delegates tasks to agents, there is an asymmetry in information and knowledge such that agents can shirk duties, distort information, and behave opportunistically. To combat these moral hazards, principals can increase monitoring of agents. A less costly approach to control agent opportunism is to align the goals of the agent to that of the principal, particularly using outcome-based contracts (Eisenhardt, 1989). Ex ante, principals can screen potential agents to mitigate adverse selection.

Problems may also emerge when agents must serve conflicting goals of multiple principals—also known as the “hydra factor” (Shapiro, 2005). In this case, the strategy of aligning agents' interests with organizational goals is confounded by conflicting goals—perhaps impossibly so. This agency problem might manifest itself in weapon system acquisition when, for instance, a program plagued by technical difficulty is jeopardized if behind schedule or over budget (threat to taxpayers' interest). Such a program could compromise the ability to deliver a system that meets end user needs (threat to end user). Additionally, jobs that are dependent on this program could be jeopardized (threat to program



executive officer's and Congress' interest). In this case, an evaluator could be biased toward a favorable past performance evaluation in order to protect the contractor and the program from scrutiny. This is an area ripe for further research (Shapiro, 2005).

In agency theory, large organizations of many people and sub-organizations are assumed to act as one homogeneous entity. This is criticized as “misplaced methodological individualism” (Worsham, Eisner, & Ringquist, 1997, p. 423). In addition to multiple principals to serve, there may be multiple evaluators (agents; Shapiro, 2005)—particularly on large, complex contracts and where performance occurs in more than one location. In cases of inter-rater disagreement, how is the principle's rating of a supplier (agent) derived?

Organizational Behavior

Contract performance often is a complex phenomenon to assess. It can involve many supplier personnel, many buyer evaluators (Wieters & Ostrom, 1979; Palmatier, 2008), multiple internal stakeholders and organizations, and multiple performance criteria at many physical locations. Often, the stakes are high such as implications to profit and future business. This is why evaluation boards for award fee (profit) determinations are comprised of many individual raters (vs. one).

Findings from organizational behavior literature are germane. Academic literature on multiple-rater performance appraisal systems (e.g., 360-degree evaluations in which superiors, subordinates, and peers evaluate the ratee) has examined the underlying premise that more raters offer more unique, valuable information about the employee's performance that would otherwise be lost if relying upon a single rater (van der Heijden & Nijhof, 2004). Additionally, more raters mitigate evaluation bias (Levy, Cawley, & Foti, 1998). While relying upon multiple ratings is thought to offer more fairness to ratees, variance in ratings is introduced attributable to individual differences in raters (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998). Thus, different raters often conclude different ratings (Dowst, 1972; Levy et al., 1998). Herein, this variance is refer to as *rating dissonance*. Given the potential for unreconciled dissonance, it is posited that

H1: There will be a negative relationship between past performance rating dissonance and past performance efficacy.

H2: There will be a positive relationship between the number of performance evaluators and past performance rating dissonance.

H3: There will be a positive relationship between the number of performance locations and past performance rating dissonance.



Performance ratings are also constrained by information flow between a rater and ratee.

Informational constraints implies that some self/supervisor discrepancies result from differing cognitions about job requirements. When performing any job, an employee must consider what tasks are to be done, how these tasks are to be performed, and what standards are to be used in judging the final outcome. Ideally, these determinations are arrived at in close consultation with the individual's supervisor, thus ensuring identical cognitions about job requirements. In reality, such complete agreement is rarely achieved. The extensive literature on role ambiguity (e.g., House & Rizzo, 1972; Jackson & Schuler, 1985; Rizzo, House, & Lirtzman, 1970) provides strong evidence that employees often do not have a clear idea of what their supervisors expect (Campbell & Lee, 1988, p. 304).

These findings are particularly relevant in service contracts where requirements are often not well defined (van der Valk & Rozemeijer, 2009). Different expectations among different performance evaluators of contractor requirements can affect performance evaluations.

Informational constraints can also stem from a supervisor's misunderstanding of the employee's job (Mitchell, 1983). Managers who are recruited from outside the company may have incomplete or inaccurate beliefs about a subordinate's job. Similarly, in situations in which jobs are highly interconnected and interdependent, a supervisor either may be unable to clearly separate the boundaries and duties of different jobs or may do so incorrectly (Kiggundu, 1981). A supervisor's misunderstanding of a subordinate's job also may reflect lack of observation (e.g., Mitchell, 1983). This has implications for a proper amount and method of monitoring suppliers. Insufficient observation can be attributed to the number of other responsibilities a manager has to the inherent nature of one's job. "Thus, it is not surprising that employees and supervisors may come to different conclusions about the employee's effectiveness. If initial cognitions about job responsibilities and standards differ, lack of agreement in ratings is inevitable" (Campbell & Lee, 1988, p. 305). Given that in contracting for services, requirements are often ill defined and given the high level of turnover in buyer-side contract administration (Hawkins et al., 2011), dissonance in supplier performance ratings should be commonplace. Buffa and Ross (2011) identified evaluator turnover as having a potential impact on supplier evaluations over time. Therefore, it is posited that



H4: There will be a negative relationship between the sufficiency of the requirement definition and past performance rating dissonance.

H5: There will be a positive relationship between the sufficiency of the requirement definition and the perceived accuracy of evaluations.

H6: There will be a positive relationship between the number of contract changes and past performance rating dissonance.

H7: There will be a negative relationship between the amount of surveillance and past performance rating dissonance.

H8: There will be a negative relationship between evaluator turnover and the perceived accuracy of evaluations.

Affective constraints also limit the amount of agreement between a supervisor's rating and ratees' self-evaluation. "If the appraisal process triggers such defense mechanisms, the end result may be described as a self-serving bias. In this context, self-serving bias refers to the tendency of individuals to take personal responsibility for successful performance, but to assign responsibility for failure to external causes" (Campbell & Lee, 1988, p. 306). In an organizational buying context, failures of a capital procurement program could be unreasonably attributed to a supplier's performance.

Sometimes the employee or the supervisor knowingly gives an inaccurate appraisal. A supervisor may do so to preserve the effectiveness of an interdependent work group (Campbell & Lee, 1988). Academic literature confirms a halo effect in employee performance appraisals (Thomas & Bretz, 1994). The same concern has specifically been raised regarding past performance evaluations (Kelman, 2010). A halo effect could partially explain inflated (i.e., inaccurate) past performance evaluations. Deliberate dishonesty is more likely to occur in self appraisals when they are used for scarce resource allocation decisions (Shrauger & Osberg, 1981). In a supplier relationship context, supplier evaluations should be accurate. Thus, it is hypothesized that

H9: There will be a negative relationship between the perceived accuracy of evaluations and past performance rating dissonance.

H10: There will be a positive relationship between the perceived accuracy of evaluations and past performance efficacy.

H11: There will be a negative relationship between a fear of a supplier dispute and the perceived accuracy of evaluations.



The acceptance of feedback affects employees' responses to feedback (Ilgen et al., 1979). "Specifically, acceptance refers to the recipient's belief that the feedback is an accurate portrayal of his or her performance" (Ilgen, Fisher, & Taylor, 1979, p. 356). This relationship was confirmed by Kinicki, Prussia, McKee-Ryan, & Wu (2004). "Previous conceptual and empirical feedback studies were based on the assumption that the specificity, frequency, and sign [positive] of feedback were independently related to the perceived accuracy of feedback" (Kinicki et al., 2004, p. 1059).

Channel Communication

In channel communication theory, Mohr and Sohi (1995) introduced the construct "distortion." Formality decreases communication distortion. Examining the government's past performance reporting system (CPARS), the reporting is quite rigid and formal. However, the collaboration between multiple raters is completely ad hoc and done outside of the CPAR system (i.e., not formal and highly variable). Therefore, it is posited that

H12: There will be a positive relationship between communication frequency and perceived accuracy of evaluations.

H13: There will be a positive relationship between communication bi-directionality and perceived accuracy of evaluations.

H14: There will be a positive relationship between communication formality and perceived accuracy of evaluations.

H15: There will be a positive relationship between past performance rating justification and past performance efficacy.

Weaknesses in evaluators' communications could be linked to resource constraints. Government acquisition personnel are overworked and, due to downsizing, understaffed. Combined, this phenomenon is referred to as *role overload*. Evaluators may simply not have sufficient time to gather the requisite facts and write thorough, sufficient justifications for past performance assessments and ratings. Likewise, evaluators may not have time to reconcile rating dissonance among multiple evaluators. Therefore, it is posited that

H16: There is a negative relationship between evaluator role overload and past performance rating justification.

H17: There is a positive relationship between evaluator role overload and past performance rating dissonance.



H18: There is a positive relationship between the perceived usefulness of a CPAR and the past performance rating justification.

H19: There is a positive relationship between the perceived accuracy of evaluations and the past performance rating justification.

Social Exchange Theory

Social exchange theory (SET) serves a prominent role in explaining exchange. SET is commonly used as a foundation for relationship marketing and buyer–seller relationships (e.g., Dwyer, Schurr, & Oh, 1987; Kingshott, 2006; Luo, 2002; Morgan & Hunt, 1994; Wilson, 1995). The foundational premises of SET may be summarized as follows. Exchange may involve both social and economic outcomes. These outcomes are compared to other exchange alternatives. Positive outcomes increase trust and commitment and, over time, norms develop that govern the relationship (Lambe, Wittmann, & Spekman 2001). Thus, SET rejects the assumption of universal opportunism and suggests that there is an alternate form of governance—the relationship. Parties to relational exchange, therefore, tend to rely more on trust, commitment, cooperation, satisfaction, and relational norms than strictly on written contracts (Heide & John, 1992).

Relational aspects have also been found to play a mediating role between suppliers' operational performance measures and a buyer's business performance. Hence, measuring performance alone does not affect business performance. Rather, measuring supplier performance increases socialization mechanisms, which, in turn, increase business performance (Cousins, Lawson, & Squire, 2008). Socialization mechanisms are structures and processes that facilitate contact between the buyers and suppliers, such as cross-functional teams, joint sessions, routine supplier conferences, and matrix reporting structures. These interactions enable each party to acquire knowledge of the others' social values and behavioral norms.

Research that developed a taxonomy of buyer–supplier relationship types (Cannon & Perreault, 1999) associated higher supplier performance evaluations to more collaborative types of relationships. Such relationships are characterized by greater operational linkages, information exchanges, cooperative norms, and buyer and supplier adaptations to each other (i.e., unique investment and customizations to processes and products for the other party's benefit). With greater channel cooperation, both intra-firm and extra-firm, it is posited that

H20: There will be a negative relationship between relationship quality and fear of a supplier dispute.



H21: There will be a negative relationship between relationship quality and past performance rating dissonance.

Returning to the first perspective of agency theory, much is said in the management, marketing, and supply chain literatures about supplier monitoring. Since increasing information via monitoring reduces uncertainty and helps prevent agent opportunism, monitoring (i.e., supplier surveillance) plays an important role in exchange relationships. As it pertains to past performance, surveillance is used to collect facts of supplier performance such as quality levels delivered, on-time performance, and generally meeting contractual requirements. These facts may be used to determine performance assessments and ratings and to bolster rating justifications. Therefore, it is posited that

H22: There will be a positive relationship between surveillance and the perceived accuracy of past performance evaluations.

H23: There will be a positive relationship between surveillance and the past performance rating justification.

Power/Dependence

Power is among the most significant phenomena in buyer–supplier relationships. It is defined as the ability to cause someone to do something that he or she would not have done otherwise (Gaski, 1984). Power emerges from its five sources: coercive, reward, expert, legitimate, and referent (French & Raven, 1959). The four sources other than coercive power were later categorized into *non-coercive* power (Hunt & Nevin, 1974). Coercive power has to do with punishments; non-coercive power coincides with rewards. The two types of power generally have opposite effects on other important constructs such as conflict and satisfaction (Gaski, 1984).

Power and dependence are two sides of the same coin (John, 1984). In government contracting, extremely high switching costs create dependence of buyers on suppliers after the award of a contract. In such cases, particularly when the buyer is less than diligent in its contract administration duties and oversight, buyers may be tempted to use past performance assessments as leverage to reap concessions from suppliers. In such cases, the accuracy of past performance evaluations could be questioned. Therefore, it is posited that

H24: There will be a positive relationship between fairness and the perceived accuracy of the past performance evaluation.

Combined, this set of propositions should explain past performance efficacy. The conceptual mode (Figure 1) is sufficiently comprehensive to enable practitioners to determine needed definitive action to improve the effectiveness of their use of past performance information.



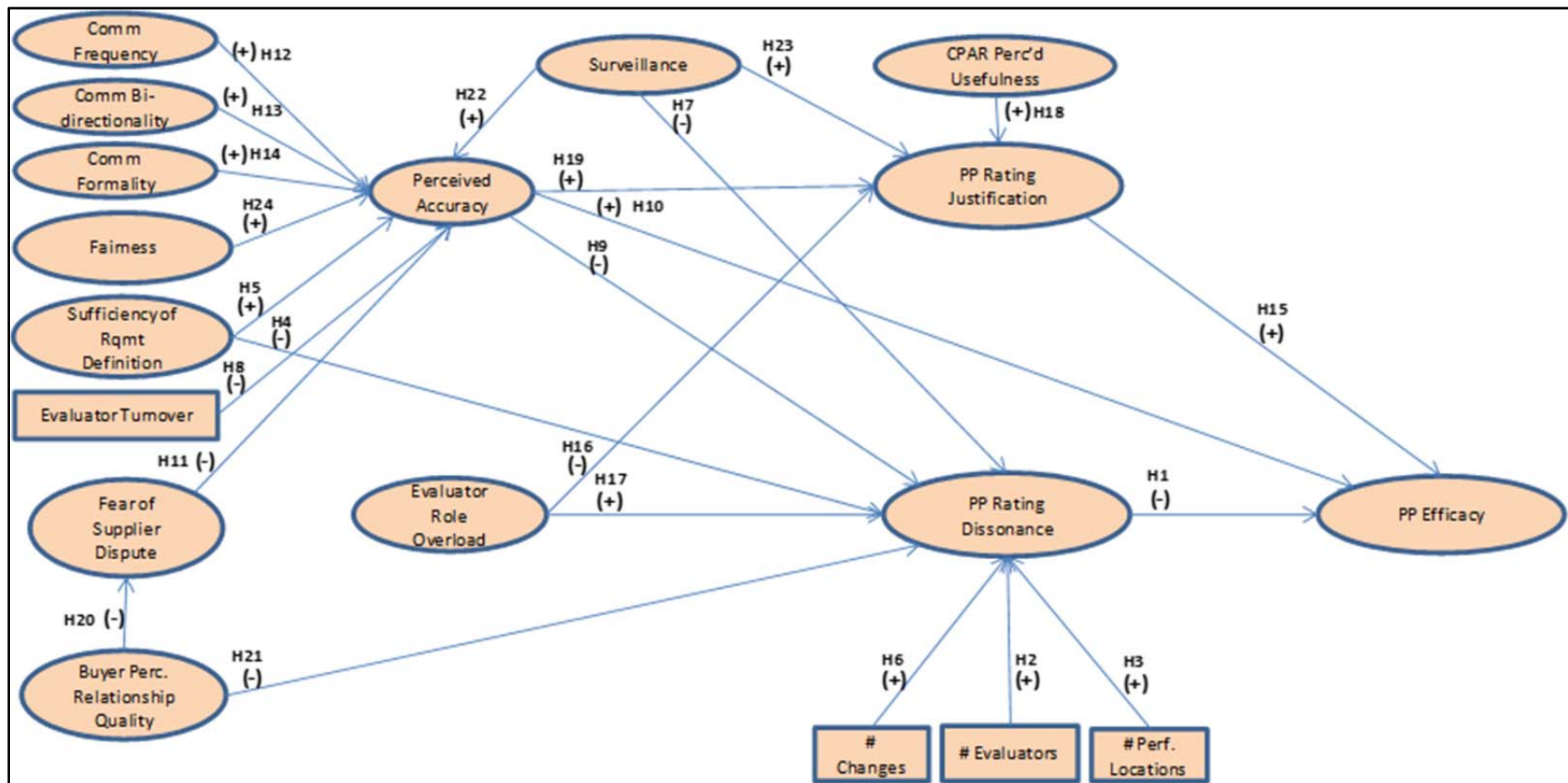


Figure 1. Conceptual Model

Note. Ovals represent latent constructs; rectangles represent objective measures.



Methodology

Qualitative Data Analysis

This research used a qualitative methodology to examine the efficacy of past performance evaluations. According to Yin (2009), a qualitative methodology is appropriate when three conditions exist: (1) The type of research question is exploratory in nature and takes the form of a “why” question, (2) the researcher has no control of the behavioral events being researched (i.e., cannot manipulate behaviors then measure results as in a controlled experiment), and (3) the focus is on contemporary events (p. 8). The research met all three criteria. Furthermore, case study research is particularly useful when researchers need to provide insight and depth to a unique phenomenon (Ellram, 1996).

Data Collection

The interview protocol (Appendix A) was developed based on a review of archival CPARs, the literature surrounding supplier performance evaluation and underlying theories discussed in the literature review, and discussions with academic experts and participants involved with past performance evaluations and source selections. In all, eight interviews were conducted. The interviews lasted between 38 and 67 minutes (mean of 51 minutes). Each interview was recorded, then transcribed. Transcripts were then sent to informants for an accuracy check, thereby enhancing construct validity (Flint, Woodruff, & Gardial, 2002; Yin, 2009). Transcripts averaged 18 pages and 7,394 words in length.

Data Analysis

The analysis process began by identifying constructs, defining those constructs, and then positing relationships between them (Patrick Van Eecke, 2006). Each interview was examined to identify themes and then tested to determine whether these themes remained consistent in subsequent interviews or in reexaminations of previous interviews. The participant interviews continued over a period of eight weeks. Initial coding led to new interviews with new participants to gain clarification and validation.

Sample

The sample of informants (Table 1) was drawn from the researcher’s personal contacts within one military service. Military and civil service employees who routinely evaluate contractor performance and enter these evaluations into the CPARS participated. These experts represented two industries that account for a large portion of the federal government’s portfolio of contract spending, aerospace and information technology (IT). Experience in evaluating contractor performance



ranged from two to 28 years, and there was a similar wide range of the number of past performance evaluations experienced (1–50). Since program managers often assume responsibility for reporting past performance evaluations into CPARS, the sample is heavily comprised of them. One contracting officer with extensive experience in CPARS, both in reporting CPARS and evaluating CPARS during source selections, was included.

Table 1. Informant Demographics

| Informant | Civilian/ Military | Industry | Years Experience | Role | Past Performance Experience (Number of Evaluations) |
|-----------|-----------------------|-----------|---------------------|------------------------|--|
| 1 | Civilian | Aerospace | 28 | Contracting Officer | 50+ |
| 2 | Military | Aerospace | 7 | Program Manager | 10 |
| 3 | Civilian | IT | 4 | Program Manager | 11 |
| 4 | Civilian | IT | 10 | Program Manager | 7 |
| 5 | Military | IT | 10 | Program Manager | 5 |
| 6 | Military | IT | 9 | Program Manager | 15 |
| 7 | Military | IT | 2 | Program Manager | 1 |
| 8 | Military | IT | 18 | Program Manager | 10 |

Results

The purpose of this research was to explain past performance efficacy. To do so, this research raised the following eight research questions:

1. Are past performance reports useful? How so, or why not?
2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?
3. Why do reviewing officials change the ratings of the evaluator (assessing official)?
4. How many man-hours does a completed past performance evaluation/rating, on average, consume?



5. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?
6. Why do past performance evaluations/ratings lack sufficient justification/supporting information?
7. Why are past performance evaluations sometimes inaccurate?
8. Do contracting officers use past performance evaluations/ratings to manage contractor performance throughout the contract, or is reporting past performance merely done to comply with the FAR (i.e., effectiveness versus compliance)?

The results of each research question are discussed in sequential order followed by excerpts from interview informants. The meanings of the excerpts are then discussed and related back to the hypothesized relationships represented in the conceptual model (Figure 1).

1. Are past performance reports useful? How so, or why not?

To examine whether past performance evaluations are seen as useful, we adopted the commonly touted utilities of past performance information. They are said to reduce performance risk in future source selections, thereby reducing contractor performance uncertainty. Past performance evaluations are also claimed to motivate contractor performance. Of the seven informants commenting on this question, the results were mixed; three agreed that past performance evaluations reduce performance risk, while four disagreed.

“I think it could be effective at mitigating a risk if the requirements that you are looking at match up with the [inaudible] past performance evaluations that you are comparing them to.”

This informant qualified a past performance evaluation as useful if it is relevant to the requirement under consideration during source selection. For source selections, relevance is a requisite criteria of past performance evaluations.

“It was a lot of fluff and I am afraid that unless everyone is really working these things to really make an impactful statement that they probably aren’t worth a whole lot if you have a lot of ones that just are fluffy.”

“Because you can’t adequately make an assessment of a contractor’s potential to perform on the future based on a ball of fluff.”

These separate informants complained that a lack of specific details hindered the utility of past performance evaluations. In other words, a lack of details can render the evaluation useless. Additionally, a lack of details can render the judgment of relevance difficult.



“I know that it is going to be watered down kind of like the [enlisted performance report/officer performance report] because there is so much pressure that the contractor puts back on the government for wording intricacies. Overall, I think I would have to question the overall overarching fairness of the process just because just like the [enlisted performance report/officer performance report] system, particularly the [officer performance report] system you question how much reality you are getting out of this if you are not seeing all of these support that goes behind the ratings. That is why I would have to say overall I would question it.”

Drawing a parallel to Air Force military personnel performance appraisals, this informant essentially commented that the past performance evaluations are inflated so as to not harm the contractor. This comment suggests support for H10, that the efficacy of a past performance evaluation could be hindered by an inaccurate (i.e., inflated) report. The next part of this comment (i.e., “pressure”) suggests a fear of a contractor’s dispute of the narrative assessments and/or ratings. In the context of the conversation, this testimony suggests support for H11, that fear of a contractor dispute may decrease the accuracy of the evaluation (i.e., rating inflation). The testimony also suggests that detailed rating justifications are needed in order to extract value (i.e., usefulness) from the past performance assessment, thus, supporting H15.

One informant commented,

“I think in concept it is not that bad. In application, it varies a lot and it is hard to get a total—the whole CPAR system is fair or not fair. I’ve seen it be fair in some places and I have seen it not be fair in some places. I have seen just a very mixed bag in a lot of places. I have seen some places and people running around with their hair on fire and it is just a task to do and they slam something out at the last second.”

This testimony infers that (1) there is variance in how past performance reports are accomplished and their quality, and (2) some assessing officials (raters) do not value the report—calling into question its utility.

Of the six informants commenting on the second part of this question, four agreed that past performance evaluations motivate contractor performance, while one informant disagreed.

“I think [a past performance evaluation] does motivate contractors to a certain extent.”

“It can be a great tool for the PM to use to motivate the contractor. I see its effectiveness on that end more so than on a source selection, if you will.”



Researcher: “So do you think—at least in your experience in those types of programs, do the CPARS tend to motivate contractors to perform?”

Informant: “I would say very minimally. It became more of an exercise of they did what they do. Then you back into these ratings and then we had a person come along different up the food chain who would review those before they went out and had different standards for what the different colors meant.”

While results were mixed as to whether past performance evaluations reduce performance risk for a future contract, most informants agreed that the evaluations motivate contractors to perform better.

2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?

Of the five informants commenting on this question, each affirmed cases in which a contract involved multiple different performance evaluators (H2). One informant commented,

“Sometimes there was some real consternation, and sometimes they actually went outside the program team and went up to higher management to get it resolved.”

The informants offered a variety of explanations for differences in assessments. Three informants mentioned different expectations of contractor performance and poor requirements definition as culprits, confirming H4 and H6 (number of changes). Two informants attributed incongruent past performance evaluations to insufficient monitoring of the contractor. This supports H7. Two informants mentioned that the different government performance evaluators had different experiences, suggesting that individual differences may exist. Two informants mentioned different locations of the contracting officer’s representative, indicating that performance may differ at different physical sites, supporting H3. Two informants also agreed that work overload precludes performance evaluators from fulfilling their duties to evaluate and document contractor performance, supporting H16 and H17.

Informant: “You have only got so many resources, and I see a number of program offices that they are doing so many things they are driving ahead of their headlights.”

Researcher: “So workload is an issue?”

Informant: “Workload is a definite.”

The following four additional reasons for dissonance among performance evaluators included a lack of facts of performance levels (H9), fear of supplier



dispute of the ratings, rater revenge, and differences in standards for ratings across evaluators (H4).

3. Why do reviewing officials change the ratings of the evaluator assessing official)?

When inquiring whether reviewing officials change ratings and/or narrative statements made by performance evaluators, the results were mixed. There appears to be plenty of opportunity for changes since several layers of management review a CPAR, as evidenced by one informant.

“From here and my boss looks at it and he is actually the program manager, [inaudible]. Then we get past them to [inaudible] one, two, three—I would say three. Three layers. If you include the contractor who eventually has a chance to look at it, that is probably a fourth layer.”

Three informants confirmed the practice, while three had no experience with changed evaluations. For those experiencing changed ratings, reasons cited included a lack of facts of contractor performance and government responsibility for contractor nonperformance.

Researcher: “You would see narrative and ratings get changed?”

Informant: “In some cases.”

Researcher: “They got changed outside of what was truly accurate or earned or deserved?”

Informant: “Many—in my opinion, many of the ratings for a long time could have been a lot lower if government had its act together and adequately supported and communicated with the contractor.”

This exchange attributes changed ratings to the government’s failure to observe or document contractor performance. The informant also mentioned a failure to communicate with the contractor.

When discussing a fear of a contractor’s dispute (i.e., a potential claim) of a past performance assessment, one offeror alluded hypothetically to diminished value of the CPAR in achieving its intended objectives. The informant then likened a change in CPARS reporting presumably to a change in a source selection rating of past performance upon being disputed (again, presumably via a bid protest). This testimony offers some evidence that a fear of a contractor’s dispute is germane to the accuracy of past performance evaluations (H11), and associated this fear to diminished past performance efficacy (H10).

“Let’s say if this gets to a legal—if we get to the point in a CPAR—and how we do a CPARS or contractor assessments—to where we are concerned and it becomes a legal fear, then I think that the value of



them will disappear and there will be no value in there. I say that from experience in a couple of situations. One of the past performance teams I led, ultimately there was a protest, and we successfully defended against the protest. The protest was denied. But, it was because we had clearly worked with legal ahead with our sections L and M in the RFP and we stuck by that methodology and we documented our methodology. And like the guy that came behind me—I deployed for a little while—and the guy that was leading the experience team took over the past performance team—he ended up spending about three days on the stand—of significant grilling. But because we had well-documented processes and we had not deviated from our section L—how we told them we were going to evaluate them—and we could substantiate them in the thing, I had no fear that we were going to [inaudible]. So we had *right* on our side, so I had no fear of standing by what we had done. In another situation I was involved in where—in process not CPARS, but very similar, to where a lot of information got watered down and changed once it became a legal matter and legal process.”

4. How many man-hours does a completed past performance evaluation/rating, on average, consume?

Six informants estimated the amount of man-hours expended to complete past performance evaluations. The quantities (8, 8, 12, 20, 80, 100) ranged from eight to 100 hours, averaging 38 hours. Larger value contracts with more complex performance monitoring and multiple sites consume more time in accomplishing a past performance evaluation in the CPARS. They also often involve greater coordination and oversight by reviewing officials. Recognizing that many different people may be involved in a single CPARS, the total effort appears to consume one man-week of effort. This represents a significant portion of time considering that performance evaluators—such as quality assurance evaluators, engineers, and contracting officer’s representatives (COR)s—often manage (i.e., report on) multiple active contracts.

5. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?

This question resembles the first part of the first research question. Question 1 inquired whether past performance evaluations are useful to reduce performance risk on future contracts. These results are less mixed, with most informants believing that past performance evaluations do not influence source selection decisions (i.e., winner determinations). One informant reported no influence. Three informants reported little influence. One informant reported some influence, and one informant reported great influence.



6. Why do past performance evaluations/ratings lack sufficient justification/supporting information?

Several informants confirmed that often past performance evaluations lack sufficient justifications for ratings and narrative assessments. In explaining why past performance evaluations lack sufficient justifications, several informants identified poor documentation of contractor performance. Poor documentation of facts could result from excess workload or a lack of contractor surveillance. Thus, support is found for H19 and H23. Two informants also identified evaluator turnover as a culprit.

“And there is a wide variety within the system, in my experience. So you get—and you find that out by calling back to the PMs that you can get ahold of, if they are still there. The older the CPARS are, obviously it is harder to find the people, and you clarify the information you are reading from a past performance perspective.”

The informant, here, referred to a high variance in quality of past performance assessments, so much so that in many cases, phone calls back to the program manager are necessary in order to validate and understand the contractor’s performance. However, this understanding is hindered by a turnover of personnel who generated the CPAR. Another informant highlighted the effect of his turnover on a CPAR.

“I was working on another project completely different from this and couldn’t even spell CPAR. I mean I didn’t really know what it was and all of a sudden I was made the program manager for a certain—for a program—and it came to, okay, it is time to do their CPAR. I wasn’t even—it was like, okay, I worked with the contractor and you know worked with the contractor to come up with what she wanted in the CPAR. Okay? At that point I was like, okay, I will write something up and send it over to them, and if it is okay with them, then we will send it forward and that was probably—I know now that is okay, you get input from them but then it is actually you writing it and then you don’t have to necessarily—you don’t have to always agree with what the contractor thinks they did. I mean sometimes you can think differently. So my first one was—and I don’t even remember what the ratings were—I really don’t, but I know that first one, that was probably—I am not going to say it was wrong, but I am going to say it was—I couldn’t have backed up some of the stuff that was in there because I wasn’t working with the contractor.”

In this case, since the informant had no experience with CPARS reporting and since, due to his recent turnover, he was not cognizant of the contractor’s performance, he essentially let the contractor write its own CPAR. Thus, support is also found for H8—that evaluator turnover diminishes the accuracy of CPARS.



7. Why are past performance evaluations sometimes inaccurate?

Informants unanimously and strongly agreed that past performance evaluations, too often, are inaccurate. Many explanations were provided by the seven informants responding to this question. Informants mentioned the following factors affecting accuracy: halo effect (unwillingness to taint a contractor's record since it could effectively lock them out of future awards), lack of facts surrounding contractor performance, inflated ratings, performance evaluator turnover (H8), differing definitions of performance standards (H5), poor requirements definition (H5), poor oversight of contractors (H22), and the disregarding of some deficiency reports.

“That is very hard to get an under satisfactory from what I have seen.”

“Many—in my opinion, many of the ratings for a long time could have been a lot lower if government had its act together and adequately supported and communicated with the contractor.”

“Some services tend to not put much negative information in there in my experience. At least the ones I have read. Some of them are written more like a performance report where it's bad to say anything negative. I think that—if that is the approach that people take, then you would take then the system has little value.”

These testimonies of separate informants confirm inflated ratings and the halo effect, which compromises accuracy. One reason underlying the inflated rating—to protect the contractor from a permanent scar—could be attributed to a concern for fairness, supporting H24. Another reason is the government's failure to observe and document contractor performance (H22).

Researcher: “To what extent do you guys worry about a dispute from a contractor or rebuttal?”

Informant: “I think the way that you address that or minimize the chance of that happening, you know, along the same lines of what these guys had said. Number one, shouldn't be any surprises on a CPAR. CPAR should not be the first time that the contractor hears about an issue. Then number two, being objective on a CPAR. If you can trace it back to your requirements or PWS and you have an objective affirmation on there, I think that reduces the chance of that happening a lot.”

This quote suggests that, consistent with H11, the fear of a contractor's dispute of the ratings or narrative assessments influences performance evaluators to collect and document supporting facts. These fact-based evaluations should improve the accuracy of the past performance evaluation.

“There were other things that were like, well, they didn't perform as well as we wanted them to, but we couldn't ding them on it because



nowhere in the contract did it specifically say this is your standard and this is where you have to meet it or exceed it.”

Researcher: “Does anybody have any experiences with accuracy—you know, issues of accuracy of the CPARS that you could tie back to something like a poorly defined requirement or not the proper amount of oversight or surveillance to the contractor?”

Informant: “We have seen a few of those things which makes the documentation part harder—or not documentation, but the supporting arguments harder, when you say, “Okay, well their requirement is this.” Well, how do you meet that because you can’t even define that?”

These quotes suggest that sometimes performance requirements are not sufficiently defined in order to collect facts and compare them to contractual requirements. Thus, support is found for H5.

“The division leadership and this particular organization has pushed down a culture that lends itself to that evidence in writing CPARS. You know the division staff pushes it down to the branch level, and the branch reviewers push that down too. So that is the first thing they look for when they are reviewing the write ups is, okay, now give me the four examples. You know if you have gone above and beyond, give me an example of that. If you have a lack of communication, give me examples of that. So that is a culture that has been pushed down to this division and that is the expectation that is displayed. The reason for that is we don’t want to go down the road for dispute. That is our defense mechanism in this particular division.”

“So we work hard in this division to have the evidence within the CPAR so it doesn’t get disputed down the road if we run into issues.”

This testimony confirms a fear of a supplier dispute, and demonstrates that this fear influences performance evaluators to bolster the justifications of their past performance ratings and narratives.

“Yes, when I was [in] the last program office that I was in, we had our support contractor, and we were meeting with that contractor virtually through email and through telephone conversation multiple times a week and constantly giving feedback. So when it was CPARS time, there were no surprises. Actually it didn’t even get disputed, and we had a couple of areas where we had a few markdowns and we had the data, and that is the important thing in writing is the data to back it up. You know, dates and documented evidence, if you will, [inaudible] come to that for an area that they may have been lacking in. So it wasn’t a surprise, just to my [inaudible] it was not a surprise for the [inaudible] contractor to get the CPAR that they did. It was constant feedback and that was just in the way of the working relationship.”



This exchange suggests an association between the buyer–supplier relationship and the quality and frequency of communications. The informant mentioned *no surprises* and *no disputes* from the contractor due to the communication. This insinuates that the evaluations were accurate and that there is, therefore, little concern for a supplier dispute. Thus, some support appears reasonable that, consistent with H20, relationship quality affects a fear of a dispute (which, in turn, affects the accuracy of the past performance evaluation).

8. Do contracting officers use past performance evaluations/ratings to manage contractor performance throughout the contract, or is reporting past performance merely done to comply with the FAR?

The general consensus to this question was negative. CPARS reporting is a one-time (or annual), ex post documentation of performance. In most cases, performance feedback is communicated via alternate communication channels and more frequently than that required by CPARS. Nonetheless, there appears to be significant variance in the rigor, frequency, quality, and amount of performance feedback across contracts. These features of communication vary by individual program managers, contract managers, CORs, or end users. This variance lends credence to H12, H13, and H14, which posit relationships between features of communication and past performance evaluation accuracy. One informant also offered testimony explicitly linking relationship quality and past performance rating dissonance (H21).

Researcher: “But it seems like you guys have a process that you use—some kind of process of communication with the contractor?”

Informant: “Right.”

Researcher: “So the team understands, the contractor understands, you talk about it throughout performance, so it sounds like it is really a nonissue.”

Informant: “Yes, I think—well, stock control I think is very unique in some ways in that the same contractor has had our contract since its inception. So we have the relationship between the functional people that we represent as a program office and the contractor in the program office. I mean, it is—I think it is an exceptional relationship there. So I think this is more of a—not a comment on CPARS, but a comment on the program.”

Researcher: “Well again, but it has a bearing and it manifests itself in the performance evaluation process.”

Informant: “Right.”

Researcher: “So I think it is a relevant factor. Sounds like it is.”

Informant: “Well, it is, yes.”



In general, qualitative interview data largely supports the conceptual model lending content validity. Informants did not specifically identify associations between past performance efficacy and evaluator dissonance (H1). Nor did they explicitly link CPAR usefulness to past performance rating justifications (H18). Most of these relationships are, however, implicit in the conversations. Explicit support discussed above was found for the remaining 21 hypotheses.

Discussion

Since its inception via the Federal Acquisition Streamlining Act of 1994 (Beausoleil, 2010), contractor past performance is intended to be an important evaluation criterion in federal source selections. The purpose was to level the playing field between the government and the contractors to mitigate information asymmetries. With more complete knowledge of contractor performance, agencies can mitigate adverse selection.

However, there are many concerns that the past performance evaluations/ratings are not properly, timely, or accurately completed. Reports often lack sufficient information to support ratings (e.g., how the contractor exceeded or failed to meet requirements) necessary to withstand a legal challenge or do not include a rating for all performance areas (OFPP, 2011). Additionally, throughout the rating process, raters often inflate ratings in order to avoid conflict with the contractor (GAO, 2009).

Unreliable or inaccurate past performance assessments can harm contractors' reputations and can bias source selections, resulting in adverse selection. If past performance information is not reliable, and if contracting officers and evaluators don't use it in discriminating between competitive proposals, the effort of collecting and reporting the past performance information is squandered. Likewise, the effort of evaluating and documenting inaccurate past performance information during source selections is wasted. Evidence suggests that the magnitude of distortion is high—so much that contracting officers, evaluators, and source selection authorities rarely use past performance information as a meaningful discriminator between proposals. In order to determine whether this seemingly vacated faith is warranted, the degree of distortion was examined.

The purpose of the research was to explore the efficacy of the government's current use of past performance information. The intent was to diagnose alleged weaknesses and to explore potential improvements. The following research questions were addressed:

1. Are past performance reports useful? How so, or why not?



2. In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why?
3. Why do reviewing officials change the ratings of the evaluator (assessing official)?
4. How many man-hours does a completed past performance evaluation/rating, on average, consume?
5. To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?
6. Why do past performance evaluations/ratings lack sufficient justification/supporting information?
7. Why are past performance evaluations sometimes inaccurate?
8. Do contracting officers use past performance evaluations/ratings to manage contractor performance throughout the contract, or is reporting past performance merely done to comply with the FAR (i.e., effectiveness versus compliance)?

This research used a qualitative methodology to examine these research questions. From a literature review, a conceptual model of 24 hypotheses was developed. Eight subject matter experts who routinely evaluate contractor performance and enter these evaluations into the CPARS were interviewed to explore the relationships posited in the model. While employing only a limited, qualitative, empirical test of the propositions, the research provides managers with some tentative guidance.

Managerial Implications

This research confirmed much of what has been reported in GAO and OFPP reports. However, the research took the next step to explain why the systemic weaknesses occur (e.g., inflated ratings, poor justifications, etc.). In doing so, several novel causal factors emerged. For example, some main findings centered around the dissonance among multiple performance evaluators on a single contract. Another major finding entailed the accuracy of evaluations and how the characteristics of channel communication play such an important role in accuracy. The findings herein introduce a plethora of implications for acquisition management, discussion of which follows.

First, dissonance across performance evaluators suggests that contractors should pay attention to evaluator dissent and develop strategies to manage each of the buyer's agents' interpretations of its performance. Government acquisition



teams and contractors might benefit from discussing during the post-award conference precisely how a situation of dissent among multiple evaluators will be managed. Additionally, since evaluator workload can affect the due diligence applied to performance evaluations, contractors could devise strategies to make the evaluators' jobs less arduous. For example, contractors can, and sometimes do, preempt the CPAR by writing their own versions of evaluations and offer these evaluations to evaluators, program managers, and contracting officers. The unintended consequence of this practice, however, is the buyer's propensity to apply less effort in its duties to independently monitor and scrutinize performance. Where buyer-supplier trust is high and where contractor performance is high and reliable, this practice of essentially outsourcing performance evaluations poses less risk. Agencies should, however, weigh the conflict of interest posed and set boundaries for this practice since it invites risk of artificial inflation of assessments.

The research also offers explanations for dissenting evaluations among multiple performance evaluators. For example, leaders should manage evaluator workload to ensure they have sufficient time to perform their past performance evaluation duties. Manning models should be more precisely developed to account for not only dollars obligated and the number of contracts awarded annually, but other time-consuming tasks such as the quantity of past performance evaluations. This research reveals that, on average, past performance evaluations consume nearly one man-week of effort. Leaders should also devise means to ensure that requirements—including measurements of success and precise definitions of CPAR ratings tailored to the requirement—are sufficiently defined prior to solicitation. These definitions should be reviewed at the post-award conference. Where interpretation can vary among evaluators, different expectations of contractor performance can emerge and fester. Likewise, the number of changes should not be excessive since this, too, can result in confusion as to what is required by the contractor, particularly on high-value, complex requirements. Inter-rater dissonance may also be reduced by ensuring that past performance assessments and ratings are more fact-based (i.e., more accurate) since it is difficult to disagree with documented facts. Finally, leaders can reduce dissonance with more proper surveillance of the contractor's work.

A central construct affecting past performance efficacy appears to be the accuracy of the evaluations. Accuracy was found to be affected by many fairly obvious factors that have been discussed in the literature, such as increased surveillance, feedback quality, bi-directionality, formality, and fear of a supplier dispute (resulting in a halo effect). These results suggest that more surveillance and performance-level measurement should be conducted in order to observe and collect the requisite facts. Thus, requiring activities should develop metrics to assess contractor performance and schedules for measurement. The results also



suggest that past performance reporting is often not a sufficient surrogate for contractor performance management. More frequent, formal, and two-way communication with the contractor is usually required, as affirmed by Steve Kelman's (2010) recommendations to improve past performance information collection and use. Thus, acquisition teams relying on the CPARS system as the sole feedback mechanism may sacrifice accuracy and, in turn, past performance efficacy. This research highlights the limitation of CPARS and a gap in federal procurement management. There is no single structured IT system and process to systematically collect, store, and synthesize contractor performance information. This is one reason why the government struggles so much to effectively manage service contracts. Supplier performance management systems are common in the for-profit sector. Examples include lasta's *SmartSupplier* scorecard tool, SAP/Ariba's *Supplier Performance Management* module, and BravoSolution's *Supplier Performance Management* tool. These structured, web-enabled tools could standardize metrics, performance data recording, analysis, and reporting. They also offer dashboard-like scorecards to assess individual contractors and groups of contractors (e.g., by commodity family or by industry). Such a structured tool could alleviate many of the weaknesses that deteriorate past performance accuracy, enable inadequate assessment justifications, and foster rater dissonance, while bolstering the government's ability to manage contractors' delivered performance levels.

In addition, several unsuspecting, novel factors emerged that explain past performance evaluation (in)accuracy. For example, informants attributed lower accuracy to evaluator turnover. This could be due to lower accountability for doing thorough work in observing and documenting contractor performance. Turnover can also exacerbate the problems caused by work overload. Thus, leaders should mitigate turnover of performance evaluators, particularly on complex contracts. Policy could also be enacted to require outgoing evaluators to conduct an interim CPAR prior to departure so that the new evaluator can begin—and assume accountability for assessing—performance evaluation at the beginning of a full evaluation period. Accuracy of past performance assessments was also affected by insufficiently defined requirements. It is difficult to assess that which is not understood or that which can have multiple interpretations. Thus, contracting officers and program managers should not move forward in contracting with ill-defined requirements. Additionally, contractors should strive to ensure that the buyer thoroughly defines requirements. An independent agency requirements ombudsman could help in this regard.

Perhaps the most novel finding is that the buyer's perceived fairness of the evaluation affects the accuracy of evaluations. This fairness can work for or against the contractor—depending on the buyer's assessment of what the contractor



deserves. On the positive side, many informants likened the one-shot, summary rating that is supposed to reflect many instances of performance to an employee's annual performance appraisal. In other words, evaluators felt it unfair to rate a contractor as below satisfactory for a single instance of a performance failure in cases where there were many other performance opportunities. Similarly, performance evaluators were reluctant to give a below satisfactory rating singularly because of the impact to the contractor's ability to secure future government business. In addition to fear of a supplier dispute to ratings, this phenomenon confirms a halo effect. Conversely, on the negative side, some performance evaluators seemed to use the past performance rating as leverage—either as a threat to a contractor during performance and prior to a CPAR or as a means to punish a contractor following poor performance (i.e., revenge). The former was particularly acute involving contracts in which the government was locked in and had little relative bargaining power compared to that of the contractor (e.g., sole source contracts).

Theoretical Implications

Agency theory has been applied to many facets of buyer–supplier exchange relationships. In this study, two dimensions of agency operate simultaneously, and a third novel dimension emerged. First, the contractor is considered an agent of the buyer in promulgating the buyer's mission. Second, the buyer (i.e., the government team) is comprised of multiple agents to itself. In the case of multiple evaluators in different organizations of the government, multiple agency relationships exist, and each can hold different interests. The third unsuspected dimension of agency pertains to the program (i.e., the requirement). In some cases, both government performance evaluators and contractor employees could begin to identify more with the program than with their employer. In other words, sometimes, what is advantageous for the program can supersede what is advantageous for either the government or the contractor. This explains the halo effect afforded a contractor who fails in one instance of performance yet the government evaluator does not mention the failure in the past performance evaluation because of a reluctance to taint the program or the contractor's chance for future business. Thus, there appears to be opportunity to examine the antecedents and consequences of quasi-agency relationships to understand under what circumstances such a quasi-agency emerges and the resultant effects.

Study Limitations

The obvious limitation of this paper is the lack of a quantitative test of the hypotheses. Thus, while serving as a foundation, future research should expand and test the propositions. These propositions lend themselves well to cross-sectional data collected via survey. The quantitative data could be analyzed using



various multivariate models such as structural equation modeling. The research also employed a limited number of interviews. While rich insights were gleaned from experienced informants, other related phenomenon may be omitted with a narrow sample.

Future Research Directions

Future research should quantitatively test the hypotheses developed herein. Such a comprehensive model with many variables and successive dependent variables could be tested via structural equation modeling. Additionally, since the scope of this study was restricted to explaining past performance efficacy (i.e., its antecedents), the consequences of an effective past performance system should be empirically explored. In other words, does a more effective past performance system result in better source selection decisions, better contractor performance, and more efficient sourcing?

Future research could also expand the context of the study. This research was constrained to the federal government sector. Research could examine the extent to which the phenomenon occurs in the for-profit sector, and could examine differences in relationships among variables attributed to the differences in sectors. Hence, is the business sector a moderator for any of the hypothesized relationships?

Future research could also delve into situations in which performance evaluators empathize with the contractor to an extent that they are willing to inflate ratings and assessments. In other cases, we see just the opposite; performance evaluators are willing to use the past performance evaluation as a sort of punishment in a vengeful way. It would be interesting to understand why different evaluators in different situations take such different approaches.

Conclusion

This research examined the long-standing systemic weaknesses in the government's processes and systems for collecting and using contractor past performance information. Major weaknesses include inaccurate evaluations and poor justifications for assessments. Consequently, often, past performance information from CPARS is not relied upon to make trade-offs in best value source selections. To explore the efficacy of past performance information collection and use, this research developed a conceptual model of key antecedents from the literature. The antecedents were then tested using interviews with subject matter experts. The testimony offers some content validity to the model, but follow-up testing should be accomplished. Factors found to affect past performance efficacy include the following: rating justification quality; contractor surveillance; multi-rater dissonance; perceived accuracy; evaluator role overload; fear of supplier dispute; perceived fairness; sufficiency of requirement definition; evaluator turnover;



relationship quality; and buyer–supplier communication frequency, bi-directionality, and formality. From these findings, important managerial and theoretical implications are drawn and future research directions are identified. It is clear that this stream of research can pay significant dividends given the substantial reliance on contractors to perform agency missions.



References

- Aberdeen Group. (2005, September). *The supplier performance measurement benchmark report*. Boston, MA: Author.
- Ashworth, R., Boyne, G. A., & Walker, R. M. (2002). Regulatory problems in the public sector: Theories and cases. *The Policy Press*, 30(2), 195–211.
- Azadegan, A. (2011). Benefiting from supplier operational innovativeness: The influence of supplier evaluations and absorptive capacity. *Journal of Supply Chain Management*, 47(2), 49–64.
- Barberis, P. (1998). The new public management and a new accountability. *Public Administration*, 76(3), 451–70.
- Beausoleil, J. W. (2010). *Past performance handbook: Applying commercial practices to federal procurement* (2nd ed.). Vienna, VA: Management Concepts.
- Benton, W. C., Jr. (2010). *Purchasing and supply chain management* (2nd ed.). New York, NY: McGraw-Hill Irwin.
- Bergen, M., Dutta, S., & Walker, O. C. (1992). Agency relationships in marketing: A review of the implications and applications of agency and related theories. *Journal of Marketing*, 56(3), 1–24.
- Buffa, F. P., & Ross, A. D. (2011). Measuring the consequences of using diverse supplier evaluation teams: A performance frontier perspective. *Journal of Business Logistics*, 32(1), 55–68.
- Burt, D. N., Dobler, D. W., & Starling, S. L. (2003). *World class supply management* (7th ed.). Boston, MA: McGraw-Hill Irwin.
- Campbell, D. J., & Lee, C. (1988). Self-appraisal in performance evaluations: Development versus evaluation. *Academy of Management Review*, 13(2), 302–14.
- Cannon, J. P., & Perreault, W. D., Jr. (1999). Buyer–seller relationships in business markets. *Journal of Marketing Research*, 36, 439–460.
- CAPS Research. (2011). *Supplier quality & delivery performance; 2011 supply management benchmarking report*. Tempe, AZ: Institute for Supply Management and W.P. Carey School of Business at Arizona State University.
- Contractor Performance Assessment Reporting System (CPARS) Metrics (n.d.). Retrieved May 19, 2011, from <http://www.cpars.csd.disa.mil/cparsfiles/cpars/metrics.htm>



- Cormican, K., & Cunningham, M. (2007). Supplier performance evaluation: Lessons from a large multinational organization. *Journal of Manufacturing Technology Management*, 18(4), 352–366.
- Cousins, P. D., Lawson, B., & Squire, B. (2008). Performance measurement in strategic buyer–supplier relationships: The mediating role of socialization mechanisms. *International Journal of Operations & Production Management*, 28(3), 238–258.
- Dowst, S. (1972, June). How designers rate suppliers—and purchasing. *Purchasing*, 12, 87–93.
- Dwyer, C. J. (2011). Linking spend analytics and contract management to public procurement. Boston, MA: Aberdeen Group.
- Dwyer, F. R., Schurr, P. H., & Oh, S. (1987, April). Developing buyer–seller relationships. *Journal of Marketing*, 51, 11–27.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Ellram, L. M. (1996). The use of the case study method in logistics research. *Journal of Business Logistics*, 17(2), 93.
- Ellram, L. M., Tate, W. L., & Billington, C. (2007). Services supply management: The next frontier for improved organizational performance. *California Management Review*, 49(4), 44–66.
- Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2012).
- Federal Procurement Data System–Next Generation (FPDS–NG). (n.d.). Retrieved from https://www.fpds.gov/fpdsng_cms/index.php/en/
- Flint, D. J., Woodruff, R. B., & Gardial, S. F. (2002). Exploring the phenomenon of customers' desired value change in a business-to-business context. *Journal of Marketing*, 66(4), 102–117.
- French, J. R. P., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150–67). Ann Arbor, MI: University of Michigan Press.
- Gaski, J. F. (1984). The theory of power and conflict in channels of distribution. *Journal of Marketing*, 48(3), 9–29.
- Giunipero, L. C., & Brewer, D. J. (1993). Performance based evaluation systems under total quality management. *International Journal of Purchasing and Materials Management*, 29(1), 35–41.



- Gordon, S. R. (2008). *Supplier evaluation and performance excellence: A guide to meaningful metrics and successful results*. Ft Lauderdale, FL: J. Ross Publishing.
- Government Accountability Office (GAO). (2001). *Trends and challenges in acquiring services* (GAO-01-753T). Washington, DC: Author.
- Government Accountability Office (GAO). (2009). *Federal contractors: Better performance information needed to support agency contract award decisions* (GAO-09-1032). Washington, DC: Author.
- Graham, J. H. (2011, January). Contractor performance assessment reports: Past performance evaluations and what to do with them. *Contract Management*, 28–35.
- Hald, K. S., & Ellegaard, C. (2011). Supplier evaluation processes: The shaping and reshaping of supplier performance. *International Journal of Operations and Production Management*, 31(8), 888–910.
- Hawkins, T., Muir, W., & Hildebrandt, G. (2011). Determinants of services sourcing performance (NPS-AM-11-C8P08R01-036). In *Proceedings of the Eighth Annual Acquisition Research Symposium*. Retrieved from <http://www.acquisitionresearch.net>
- Heide, J. B., & John, G. (1992). Do norms matter in marketing relationships? *Journal of Marketing*, 56(4), 32–44.
- House, R., & Rizzo, J. (1972). Role conflict and ambiguity as critical variables in a model of organizational behavior. *Organizational Behavior and Human Performance*, 7(3), 467–505.
- Hunt, S. D., & Nevin, J. R. (1974, May). Power in a channel of distribution: Sources and consequences. *Journal of Marketing Research*, 11, 186–93.
- Husted, C., & Reinecke, N. (2009, Summer). Improving public sector purchasing. *McKinsey Quarterly*, 4, 18–25.
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371.
- Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Performance*, 36, 16–78.
- John, G. (1984). An empirical investigation of some antecedents of opportunism in a marketing channel. *Journal of Marketing Research*, 21(3), 278–289.



- Kelman, S. (2010). Three ways to revitalize the use of past-performance data. *Federal Computer Week*. Retrieved from <http://fcw.com/blogs/lectern/2010/05/steve-kelman-past-performance-source-selection.aspx>
- Kern, D., Moser, R., Sundaresan, N., & Hartmann, E. (2011). Purchasing competence: A stakeholder-based framework for chief purchasing officers. *Journal of Business Logistics*, 32(2), 122–138.
- Kiggundu, M. N. (1981). Task interdependence and the theory of job design. *Academy of Management Review*, 6, 499–508.
- Kingshott, R. P. J. (2006). The impact of psychological contracts upon trust and commitment within supplier–buyer relationships: A social exchange view. *Industrial Marketing Management*, 35(6), 724–739.
- Kinicki, A. J., Prussia, G. E., McKee-Ryan, F. M., & Wu, B. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology*, 89(6), 1057–1069.
- Lambe, J. C., Wittmann, C. M., & Spekman, R. E. (2001). Social exchange theory and research on business-to-business relational exchange. *Journal of Business-to-Business Marketing*, 8(3), 1–36.
- Leenders, M. R., Johnson, P. F., Flynn, A. E., & Fearon, H. E. (2006). *Purchasing and supply management* (13th ed.). Boston, MA: McGraw-Hill Irwin.
- Levy, P. E., Cawley, B. D., & Foti, R. J. (1998). Reactions to appraisal discrepancies: Performance ratings and attributions. *Journal of Business and Psychology*, 12(4), 437–455.
- Limberakis, C. G. (2011). *The year of the supplier*. Boston, MA: Aberdeen Group.
- Lord, R. D. (2005). *Contractor past performance information (PPI) in source selection: A comparison study of public and private sector* (NPS-CM-04-019). Retrieved from <http://www.acquisitionresearch.net>
- Luo, X. (2002). Trust production and privacy concerns on the Internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management*, 31(2), 111–118.
- Minihan, T. (2007). The business case for supplier performance. *Supply Excellence*. Retrieved from <http://supplyexcellence.com/blog/2007/04/18/the-business-case-for-supplier-performance-management>
- Mitchell, T. R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F. Landy, S. Zedeck, & J.



- Cleveland (Eds.), *Performance measurement & theory* (pp. 39–59). Hillsdale, NJ: Erlbaum.
- Mohr, J. J., & Sohi, R. S. (1995). Communication flows in distribution channels: Impact on assessments of communication quality and satisfaction. *Journal of Retailing*, 71(4), 393–416.
- Monczka, R. M., Choi, T. Y., Kim, Y., & McDowell, C. (2011). Supplier relationship management: An implementation framework. Tempe, AZ: CAPS Research.
- Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2011). *Purchasing and supply chain management* (5th ed.). Mason, OH: South-Western Cengage Learning.
- Morgan, R. M., & Hunt, S. D. (1994, July). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58, 20–38.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557–576.
- Niskanen, W. A. (1968). The peculiar economics of bureaucracy. *American Economic Review*, 58(2), 293–305.
- Office of Federal Procurement Policy (OFPP). (2011, January 21). *Improving contractor past performance assessments: Summary of the Office of Federal Procurement Policy's review, and strategies for improvement* [Memorandum]. Washington, DC: Office of Management and Budget.
- Palmatier, R. W. (2008, July). Interfirm relational drivers of customer value. *Journal of Marketing*, 72(4), 76–89.
- Patrick Van Eecke, G. S., Freund, W., Goeskjaer, M., & Ooms, B. (2006). *Legal study on unfair commercial practices within b2b e-markets: Final report* (European Commission Study ENTR/04/69). Retrieved from http://www.pedz.uni-mannheim.de/daten/edz-h/gdb/06/b2b_2006_ls.pdf
- Perkins, W. S. (1993). Measuring customer satisfaction. *Industrial Marketing Management*, 22(3), 247–254.
- Prahinski, C., & Benton, W. C. (2004). Supplier evaluations: Communication strategies to improve supplier performance. *Journal of Operations Management*, 22, 39–62.
- Prahinski, C., & Fan, Y. (2007). Supplier evaluations: The role of communication quality. *Journal of Supply Chain Management*, 43(3), 16–28.



- Purdy, L., & Safayeni, F. (2000). Strategies for supplier evaluation: A framework for potential advantages and limitations. *IEEE Transactions on Engineering Management*, 47(4), 435–443.
- Reilly, S. (2012). How the Air Force blew \$1 billion on a dud system. *Federal Times.com*. Retrieved from <http://www.federaltimes.com/article/20121126/DEPARTMENTS01/311260009/>
- Rizzo, J., House, R., & Lirtzman, S. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly*, 15, 150–163.
- Rudzki, R. A., Smock, D. A., Katzorke, M., & Stewart, S., Jr. (2006). *Straight to the bottom line: An executive's roadmap to world class supply management*. Ft Lauderdale, FL: J. Ross Publishing.
- Schmitz, J., & Platts, K. W. (2003). Roles of supplier performance measurement: Indication from a study in the automotive industry. *Management Decision*, 41(8), 711–721.
- Shapiro, S. P. (2005). Agency theory. *Annual Review of Sociology*, 31, 263–284.
- Shrauger, J., & Osberg, T. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin*, 90, 322–351.
- Simpson, P. M., Siguaw, J. A., & White, S. C. (2002, Winter). Measuring the performance of suppliers: An analysis of evaluation processes. *Journal of Supply Chain Management*, 38(4), 29–41.
- Talluri, S., & Sarkis, J. (2002). A model for performance monitoring of suppliers. *International Journal of Production Research*, 40(16), 4257–4269.
- Thomas, S. L., & Bretz, R. D. (1994). Research and practice in performance appraisal: Evaluating employee performance in America's largest companies. *SAM Advanced Management Journal*, 59(2), 28–34.
- Trent, R. J. (2007). *Strategic supply management: Creating the next source of competitive advantage*. Ft Lauderdale, FL: J. Ross Publishing.
- van der Heijden, B. I. J. M., & Nijhof, A. H. J. (2004). The value of subjectivity: Problems and prospects for 360-degree appraisal systems. *International Journal of Human Resource Management*, 15(3), 493–511.
- Van der Valk, W., & Rozemeijer, F. (2009). Buying business services: Towards a structured service purchasing process. *Journal of Services Marketing*, 23(1), 3–10.



- Wathne, K. H., & Heide, J. B. (2000). Opportunism in interfirm relationships: Forms, outcomes, and solutions. *Journal of Marketing*, 64(4), 36–51.
- Wieters, D. C., & Ostrom, L. L. (1979). Supplier evaluation as a new marketing tool. *Industrial Marketing Management*, 8(2), 161–166.
- Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York, NY: The Free Press.
- Wilson, D. T. (1995). An integrated model of buyer–seller relationships. *Journal of the Academy of Marketing Science*, 23(4), 335–345.
- Worsham, J., Eisner, M. A., & Ringquist, E. J. (1997). Assessing the assumptions: A critical analysis of agency theory. *Administration and Society*, 28(4), 419–440.
- Yin, R. K. (2009). *Case study research: Design and methods*. Los Angeles, CA: Sage Publications.



Appendix A. Interview Questionnaire

How should CPARS assist in future source selections?

Are CPARS reports/PP evaluations useful to buyers/PMs? How so, or why not?

- Do they reduce future supplier performance uncertainty?
- What evidence have you seen/heard that suggests that past performance evaluations motivate suppliers to perform?

What are the consequences of not being able to reduce future supplier performance uncertainty?

What factors affect the ability to reduce future supplier performance uncertainty?

To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions?

In the cases of multiple evaluators on a single contract (or TO/DO), to what extent do past performance evaluations/ratings deviate among evaluators?

In the cases of multiple evaluators on a single contract (or TO/DO), why do evaluations differ among evaluators?

of evaluators?

Differing objectives/interests by multiple evaluators?

of changes/modifications?

Insufficient monitoring?

Lack of facts of performance level?

Poor requirements definition?

Poor relationship with contractor?

Fear of a dispute to ratings?

Too much workload – not willing/able to endure the anticipated conflict/rebuttal?

Complexity of the requirement (multiple locations, uncertainty, dollar value, performance risk)?

In the cases of multiple evaluators on a single contract (or TO/DO), what processes do evaluators use to reduce evaluations to one?

Why do Reviewing Officials change the ratings of the evaluator (Assessing Official)?

To what extent do past performance evaluations/ratings deviate from the contractor's assessment?



Why is there a difference between the initial past performance evaluation/rating and that of the contractor?

How many man-hours does a past performance evaluation/rating, on average, consume?

Why do past performance evaluations/ratings lack sufficient justification/ supporting information?

Not enough time or too much workload?

Insufficient monitoring?

Lack of facts of performance level?

Not seen as useful?

Do contracting officers use past performance evaluations/ratings to manage contractor performance throughout the contract?

How is contractor performance actively managed during (vs. after) performance?

Are past performance evaluations accurate? If not, why not?

Poor requirements definition?

Insufficient monitoring?

Evaluator turnover?

To what extent are ratings inflated?

Why do buying organizations change their past performance ratings?

Fear of a dispute?

Poor rating justification?

Insufficient monitoring?

Lack of facts of performance level?

Poor requirements definition?

Do you suspect that the government ever uses the past performance rating/evaluation as leverage? If so, how?

Do you suspect that the government ever uses the past performance rating/evaluation as a means of achieving justice (i.e., ensuring that the contractor gets what it deserves)?

Is the past performance evaluation process fair?

What else should I consider about the past performance process and/or outcomes?

Demographics:



How many past performance evaluations have you participated in?

Duty title: _____

Years of experience?





ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net