WKU-CM-16-150


# ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

## Antecedents and Consequences of Supplier Performance Evaluation Efficacy

30 June 2016

**Dr. Timothy G. Hawkins, Assistant Professor**

Gordon Ford College of Business

**Western Kentucky University**


"This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-15-1-0057, awarded by the Naval Supply Systems Command Fleet Logistics Center San Diego (NAVSUP FLCSD). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government."


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

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

Supplier performance evaluation (SPE) 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). Numerous weaknesses associated with industrial buyers' collection and use of SPE information (a.k.a., past performance information) have been documented in the for-profit and not-for-profit sectors. These weaknesses call into question the efficacy of SPEs. Neither the factors affecting SPE efficacy (i.e., its antecedents) nor the effects of SPE efficacy (i.e., consequences) on suppliers have been empirically explored. Despite the fallibility of SPE schemes, there are no known studies that explore the accuracy of SPEs, nor are there studies examining whether and how inaccurate SPEs affect suppliers – specifically, their performance. The purpose of this research, therefore, is to identify the factors affecting SPE efficacy, then to examine how SPE efficacy, in turn, affects supplier outcomes. This research employs a mixed method of qualitative interviews and quantitative analysis of survey data collected from suppliers and from assessors of supplier performance. Based on the findings, the research makes several contributions to theory and practice, and provides directions for future research.

**Keywords:** Supplier Performance Evaluation, Past Performance, Contract Management, Supplier Reputation, Transaction Costs, Adverse Selection, Rating Justification, Rating Dissonance

THIS PAGE LEFT INTENTIONALLY BLANK

# About the Author

**Timothy G. Hawkins, Lt Col (Ret.), USAF, PhD**, is an assistant professor in the Department of Marketing at Western Kentucky University. He researches and teaches in the realm of supply chain management, government contracting, and strategic sourcing. He has 20 years of sourcing experience in industry and government. Dr. 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 Logistics*, *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*, *Journal of Contract Management*, *International Journal of Procurement Management, Journal of Product and Brand Management,* 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 (Ret.), USAF, PhD
Department of Marketing
Western Kentucky University
1906 College Heights Blvd, #21059
Bowling Green, KY 42101
Tel: (270) 745-2412
E-mail: timothy.hawkins@wku.edu

THIS PAGE INTENTIONALLY LEFT BLANK

WKU-CM-16-150


# ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

## Antecedents and Consequences of Supplier Performance Evaluation Efficacy

30 June 2016

**Dr. Timothy G. Hawkins, Assistant Professor**

Gordon Ford College of Business

**Western Kentucky University**


"This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-15-1-0057, awarded by the Naval Supply Systems Command Fleet Logistics Center San Diego (NAVSUP FLCSD). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government."

THIS PAGE INTENTIONALLY LEFT BLANK

# Table of Contents

# List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

# List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

# Introduction

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 (i.e., the buying organization). Substantial transaction costs are dedicated to avoid *adverse selection* — the risk of selecting an incapable supplier that otherwise misrepresents itself as capable. Following contract formation, more transaction costs are incurred to monitor supplier performance to thwart supplier opportunism ex post.

Supplier performance evaluation (SPE) became popular in the 1950s (Wieters and Ostrom,1979), and now SPE is an essential best practice in business-to-business sourcing (Gordon, 2008; Talluri and Sarkis, 2002). SPE 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). SPEs 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 and Platt, 2003). Specifically, SPEs inform source selection decisions of the likelihood that a prospective supplier will successfully perform the contract (FASA, 1994).

Similarly, the primary purpose of the U.S. federal government's Contractor Performance Assessment Reporting System (CPARS) "is to ensure that current, complete and accurate information on contractor performance is available for use in procurement source selections" (Naval Sea Logistics Center Portsmouth, 2014, p.1). The idea is that by better informing source selection decisions, better best value selections will occur. Integrally related is the supplier's level of performance. If performance levels are assessed and recorded, and if this information is available to buyers during a future source selection, it is believed that suppliers will work harder to ensure satisfactory (or better) performance (OFPP, 2000).

Despite long-term awareness of weaknesses and despite recent, concerted, high-level efforts to improve past performance reporting, the government's past performance evaluations of its suppliers continue to be deficient (GAO, 2014). Too often, they 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 are often inclined to 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 buyers 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 efforts of prospective suppliers in documenting and of buyer-side evaluators in evaluating inaccurate past performance information during source selections is wasted. Notably, we don't know how much transaction costs by all parties involved are 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. Notwithstanding, buying organizations often use SPE information to identify and rank superior performing suppliers. Of course, the rankings and status are suspect if the underlying SPE ratings are not accurate.

Problems are not unique to the not-for-profit sector. Hald and Ellegaard (2011) found that supplier evaluations change throughout the evaluation process. Underlying data captured in enterprise resource planning (ERP) databases is often flawed. Masses of performance data are condensed into more general ratings sacrificing fidelity. Buyers also commonly use multiple evaluators to rate supplier performance (Hald and Ellegaard, 2011; Buffa and Ross, 2011), which invites different perspectives of supplier performance. To what extent does the evaluators' dissonance affect perceived accuracy of SPEs? 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.

Despite the fallibility of SPE schemes, there are no known studies that explore the accuracy of SPEs. Therefore, further investigation is needed in order to explore the validity of SPE processes. After all, SPE assessments can affect key outcomes such as contract compliance, supplier performance payments, supplier reputation, future business awards, incentive awards, and status achievement (e.g., a "preferred" supplier). As such, the effectiveness of SPEs in assisting source selection decisions is questionable (Berrios, 2006). In other words, we do not know the extent to which SPEs validly build the buyer's confidence in its assessment of the risk of doing business with a particular

supplier prior to contract award. Furthermore, the impact of deficient SPEs on the industrial supply base is unknown.

## Scope and Objectives

The purpose of this research, therefore, is to explain the efficacy of SPE and to explore the effects of SPE efficacy on supplier outcomes such as performance and relationship quality. This research explores the extent to which the supplier performance information collection and usage processes achieve the intended goals of: (1) mitigating the risk of adverse selection, and (2) motivating supplier performance. The following research questions are explored:

1. What factors decrease the efficacy of SPEs?
2. How do suppliers react to inaccurate SPEs?
3. Do SPEs, in general, motivate suppliers to increase performance?
4. How does the accuracy of SPEs affect relationship quality?
5. Why are SPEs often inaccurate?
6. How many man-hours do suppliers invest in responding to SPEs?
7. What communication tactics do suppliers use to manage the SPE process?
8. To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy?

The answers to these eight questions should help identify the antecedents and consequences of SPE. The remainder of this paper is organized in the following manner. The research explores antecedents and consequences to SPE efficacy, and uses two separate approaches to do so. To explore the antecedents, this research builds off of prior research (Hawkins, 2013) to test previously-suggested propositions of buyer-side factors that affect SPE efficacy. To identify the consequences of SPE efficacy on suppliers, an exploratory, qualitative approach is employed. Likewise, the research is organized in this split manner. The first part of the methodology and results section addresses the antecedents in a quantitative, hypothesis testing, confirmatory approach. In contrast, the second part is exploratory, seeking to identify outcomes of SPE efficacy on suppliers. First, a literature review is presented describing the conceptual framework and hypotheses. Next, the study presents the research designs and methodologies. Lastly, discussion, limitations, implications, future research directions, and conclusions are offered.

THIS PAGE LEFT INTENTIONALLY BLANK

# Literature Review

Similar to the conclusion of Ashworth et al. (2002), a single comprehensive theoretical framework explaining the efficacy of collecting and using supplier past performance information was not found. 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 the government's past performance policies and SPEs in general.

## Past Performance

U.S. federal government contracting serves as the context for this study due to its expansive scope (dollars, industries, and geographies), rigor, established fairness, and standardized procedures. 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. By necessity, 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, 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 to the government, deficient oversight of contractors, deficient contract administration, and the government's lack of contractor performance management (GAO, 2009).

It is also important to note the U.S. Military Departments' recently-emerged practice of ranking government contractors based on performance across multiple contracts. This annual ranking, deemed the superior supplier incentive program (SSIP), relies on performance data from CPARs (USD AT&L, 2015). The purpose is to incentivize contractor performance, and to recognize those top achievers. The SSIP ranks the top 30 suppliers defined by the highest 3-year dollar obligations, and ranks the suppliers' business units at the business unit level. Suppliers deemed a *superior supplier*, are eligible for relaxed or more favorable contract terms and conditions (e.g., progress payment retention percentage, increased intervals between business system reviews, priority for adjudicating final labor and indirect cost rates, etc.). Hence, the efficacy of the SPE process takes on additional meaning by providing firms bragging rights (i.e., marketing material and enhanced reputation) and eased admininstrative burdens.

## Supplier Performance Evaluation

Supplier performance management (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 et al., 2011b). 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 et al., 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 65.5 % reduction in the number of suppliers and a 61.5% 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 Research (Monczka et al., 2011a) 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 et al., 2003; Leenders et al., 2006; Monczka et al., 2011; Rudzki et al., 2006; Trent, 2007). Not surprisingly, SPM is a core competence of chief procurement officers (Kern et al., 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 enterprise resource planning (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).

Given the above findings, the focal outcome of interest of this study is *SPE Efficacy* – defined herein as the extent to which SPEs achieve the two stated goals of motivating supplier performance and, during source selection, mitigating the risk of unsuccessful performance (i.e., avoid adverse selection). The ensuing review of the relevant literature identifies the central factors affecting SPE efficacy, then peels the onion back further to unveil their antecedents.

## 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 organization, their local organization or unit, and external stakeholders (e.g., shareholders or taxpayers in the public sector). Problems of agency arise when agents' self-interests differ from his or her employer's goals (Bergen et al., 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, evaluators often fail to properly monitor a supplier's performance. If the supplier's performance did not meet requirements, rather than rate the supplier 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 SPE in order to protect the supplier 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 et al., 1997, p. 423). In addition to multiple principals to serve, there may be multiple evaluators (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? Given

these problems of agency, *rating dissonance* is among the central constructs of this study. The variance in ratings due to multiple evaluators of supplier performance is referred to herein as *rating dissonance*.

## 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.

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 et al., 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 et al., 1998). Thus, different raters often conclude different ratings (Dowst, 1972; Levy et al., 1998), which may be attributed to different backgrounds, observing different instances of supplier performance, and different interpretations of the meaning of performance critieria and rating definitions. These differences take time and effort to resolve and internally agree upon a single rating or narrative.

Multiple raters may be indicators of complexity (e.g., multiple points of failure and multiple locations). Suppliers may be able to more successfully rebut ratings under high complexity. Suppliers may also be more able to offset relatively minor failures with their successes, garnering an overall rating that is acceptable to the supplier. If a supplier can "escape" unscathed in the rating (i.e., no threat), there is little need to increase performance, and little threat of negative performance information being discovered during a future source selection. Given the potential for unreconciled dissonance, it is posited that:

**H1: There will be a negative relationship between rating dissonance and SPE efficacy.**

**H2: Rating dissonance will be positively related to the number of hours to complete the SPE.**

**H3: The lower the accuracy, the greater the number of hours to complete the SPE.**

In federal government contracting, suppliers are provided the SPE ratings and given an opportunity to respond, rebut, agree and otherwise comment. Disagreements are elevated to a reviewing official at least a level above the assessing official for resolution. Resolution takes effort expended to explain original positions internally and to seek the facts substantiating the ratings. Thus, supplier disputes, while allowed, are not necessarily welcomed. This phenomenon is not unique to government contracting; suppliers to for-profit businesses may have executive-level relationships within the buying organization and may use those communication channels to voice disagreement with SPEs. Herein, this phenomenon is defined as fear of a supplier dispute. Attempts among multiple raters to thwart a supplier rebuttal may invite internal conflict. Some evaluators may be inclined to inflate ratings to avoid a dispute, while others may take a legalistic, strict approach. If inflated, accuracy suffers. Given the above logic, it is hypothesized that:

**H4: The lower the perceived accuracy, the greater the fear of supplier dispute.**

**H5: There will be a positive relationship between fear of supplier dispute and 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 et al., 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, is it posited that:

**H6: There will be a negative relationship between the sufficiency of the requirement definition and rating dissonance.**

**H7: There will be a positive relationship between the sufficiency of the requirement definition and perceived accuracy.**

**H8: There will be a negative relationship between evaluator turnover and perceived accuracy.**

*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 SPEs (Kelman, 2010). A halo effect could partially explain inflated (i.e., inaccurate) SPEs.

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 may also be tainted by a supplier seeking to preserve its reputation. Suppliers may refute any negative information being recorded regardless of its accuracy. To do so, they often challenge the rating and/or justification, which causes more effort by the buying organization to resolve disagreements. If buying organizations either can't muster the evidence to justify a particular rating and/or consciously decide not to bother with the trouble to debate the rating, accuracy can suffer. Thus, it is hypothesized that:

**H9: There will be a negative relationship between perceived accuracy and rating dissonance.**

**H10: There will be a positive relationship between perceived accuracy and SPE efficacy.**

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 et al., 1979, p. 356). This relationship was confirmed by Kinicki et al. (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 "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 occurs outside of the CPAR system (i.e., not formal and highly variable). In examining channel communication, often three aspects of communication are explored – formality, bi-directionality, and frequency. If these three facets of communication among exchange members increases, more information is shared, better understandings are attained, and therefore, the accuracy of SPEs should increase. Therefore, it is posited that:

**H11: There will be a positive relationship between communication frequency and perceived accuracy.**

**H12: There will be a positive relationship between communication bi-directionality and perceived accuracy.**

**H13: There will be a positive relationship between communication formality and perceived accuracy.**

Weaknesses in evaluators' communications could be linked to resource constraints. Government acquisition personnel are often overworked and 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 SPE assessments and ratings. Likewise, evaluators may not have time to reconcile rating dissonance among multiple evaluators. Therefore, it is posited that:

**H14: There is a negative relationship between role overload and rating justification.**

**H15: There is a positive relationship between role overload and rating dissonance.**

Critics contend that SPEs are often not accurate, and therefore the SPE system (e.g., CPARS) is not useful. If not factual and detailed, the SPEs cannot motivate suppliers to work harder and cannot provide insights that reduce the risk of adverse selection in the future. Hence, absent accuracy, SPEs become less useful. Further, if the SPE scheme is not useful, evaluators will not put forth the effort required to develop a detailed, factual rating justification that will be accepted by the supplier and, if rebutted, internally by the reviewing official., Thus, it is posited that:

**H16: There is a positive relationship between perceived usefulness and rating justification.**

**H17: There is a positive relationship between perceived accuracy and rating justification.**

**H18: There will be a positive relationship between rating justification and SPE efficacy.**

**H19: There will be a positive relationship between perceived accuracy and perceived usefulness.**

## 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 et al., 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). Contracts are incomplete, and can be costly and inefficient to administer as their details increase. Relational exchange renders the exchange more efficient.

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. Interactions entail communications. Communication increases trust (Morgan and Hunt, 1994), a central construct to effective relational exchange.

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: Communication frequency will be positively related to relationship quality.**

**H22: Communication bi-directionality will be positively related to relationship quality.**

**H23: There will be a positive relationship between communication formality and relationship quality.**

**H24: Turnover will be negatively related to relationship quality.**

Returning to 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 SPEs, 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 ratings. Therefore, it is posited that:

**H25: There will be a positive relationship between surveillance and perceived accuracy.**

One relational norm important to effective exchange is fairness (Kumar et al., 1995). Often the concept is referred to as distributive justice, referring to the extent to which each exchange member's cost-benefit ratios are approximately equal. Government buyers in particular have a duty to treat suppliers fairly. In the for-profit sector, fair treatment of suppliers is paramount to effective relationship quality (Kumar et al., 1995). In an SPE context, fairness pertains to the extent to which the supplier is given the performance ratings it deserves (i.e., that which it earned). Fair ratings are those that have been earned, no more and no less. Particularly in cases in which requirements are not well defined, the criteria for evaluating supplier performance are not well defined, and/or the ratings used to assess performance are not well defined (or invite wide latitude in interpretation), a supplier must rely on the buyer to be fair. A deviation from a fair rating would insinuate a rating that is not right – or less than accurate.

**H26: There will be a positive relationship between fairness and perceived accuracy.**

## 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 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. Additionally, sole source contracts are commonplace

which gives rise to buyer dependence (and supplier power). In such cases, particularly when the buyer is less than diligent in its contract administration duties and oversight, buyers may be tempted to use SPEs as leverage to reap concessions from suppliers. In cases where ratings are subtly bargained for some concession, the accuracy of SPEs could be questioned. Therefore, it is posited that:

**H27: Leverage attitude will be negatively related to perceived accuracy.**

Combined, this set of propositions should explain SPE 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 SPEs.

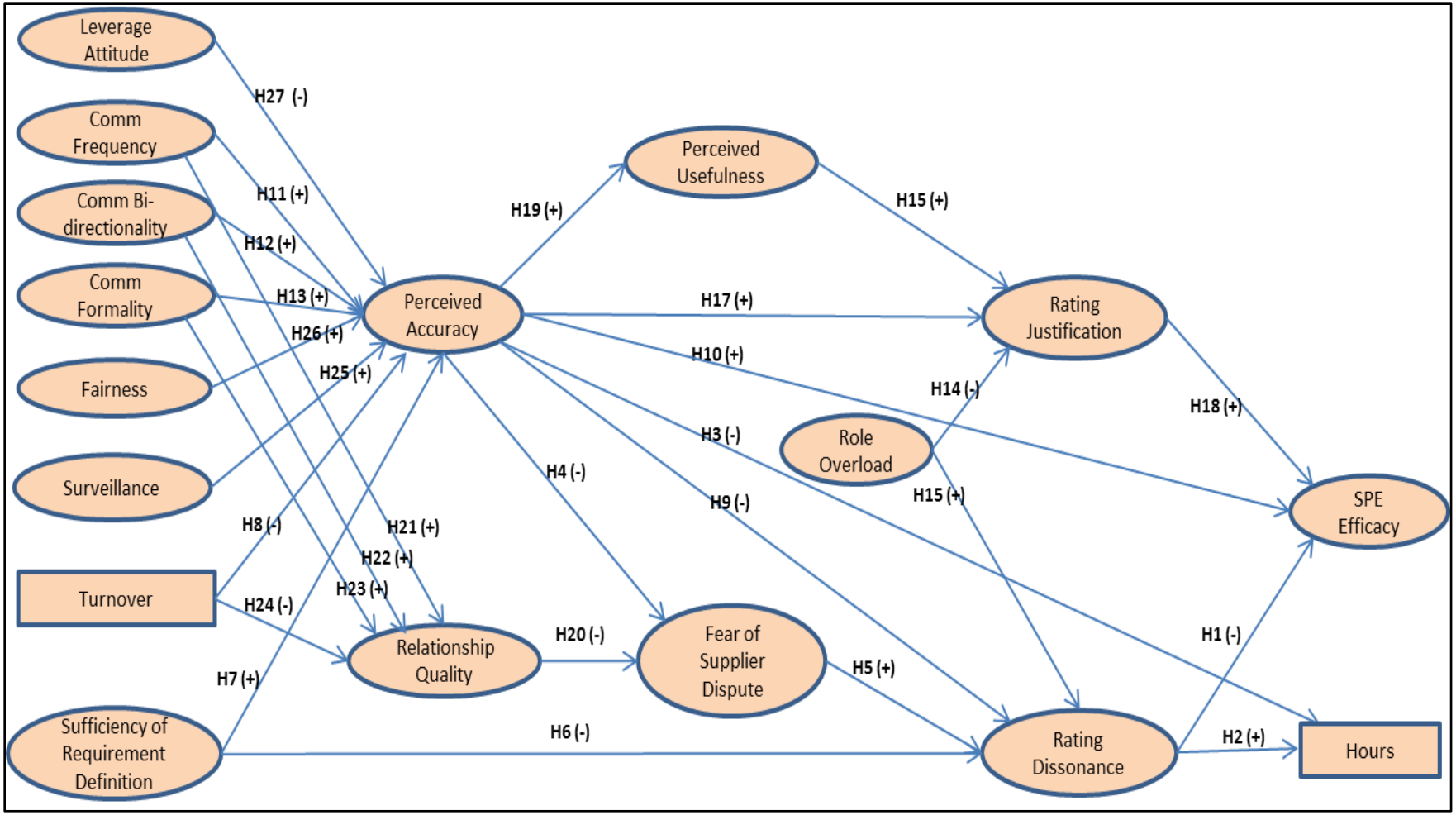


Figure 1. Conceptual Model

*Note:* Ovals represent latent constructs; rectangle represents objective measure

THIS PAGE LEFT INTENTIONALLY BLANK

# Methodology

The purpose of this research is to explain the efficacy of SPE and to explore the effects of SPE efficacy on supplier outcomes such as performance and relationship quality. This research explores the extent to which the supplier performance information collection and usage processes achieve the intended goals of: (1) mitigating the risk of adverse selection, and (2) motivating supplier performance. Table 1 lists the eight research questions that were explored, and indicates the research method and object for each. This research employed quantitative and qualitative methodologies to examine the antecedents and consequences of supplier performance evaluation efficacy. First, the quantitative methodology and results are detailed, and then the qualitative procedures and results are described.

Table 1. Research Questions

| No. | Research Question | *Research Object | **Research Method |
|---|---|---|---|
| 1 | What factors decrease the efficacy of SPEs? | B & S | Qt & Ql |
| 2 | How do suppliers react to inaccurate SPEs? | S | Ql |
| 3 | Do SPEs, in general, motivate suppliers to increase performance? | S | Ql |
| 4 | How does the accuracy of SPEs affect relationship quality? | B & S | Qt & Ql |
| 5 | Why are SPEs often inaccurate? | B & S | Qt & Ql |
| 6 | How many man-hours do suppliers invest in responding to SPEs? | S | Ql |
| 7 | What communication tactics do suppliers use to manage the SPE process? | S | Ql |
| 8 | To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy? | B | Qt |

*B=buyer; S=supplier
**Qt=Quantitative; Ql=Qualitative

## Quantitative Data Analysis

This research examines the antecedents and consequences of supplier performance evaluation. Thus, the most appropriate unit of analysis was the transaction (i.e., the contract or delivery/task order). The most appropriate individual to provide data on contractor performance, the respective contract, and the situation was the CPARS Assessing Official. In order to mitigate self-selection response bias, respondents were asked to complete the survey with respect to their most recently completed CPAR.

## Measurement

The model included objective variables and latent constructs. Existing scales with established reliability and validity were used to measure latent constructs. For new constructs with no existing scales, measures were created from interviews with assessing officials. The following latent constructs were measured with newly created scales: past performance efficacy, past performance rating dissonance, rating justification, leverage attitude, and fear of supplier dispute. These scales were developed from in-depth interviews with eight performance-assessing officials. For details of the interview methodology and informant demographics, see Hawkins (2013).

Surveillance was measured using a four-item scale from Stump and Heide (1996). Communication formality was measured using a five-item scale from Prahinski and Benton (2004). The role overload construct was measured with four items from House and Rizzo (1972). Perceived usefulness was measured with a six-item scale adapted from Davis (1989). Relationship quality assessed satisfaction, trust, and commitment as developed by Palmatier (2008). Commitment was measured using a four-item scale developed by Kumar et al., (1995). Trust was measured with a six-item scale from Johnson et al., (2004). Satisfaction was measured using a five-item scale from Cannon and Perault (1999). A five-item scale adapted from Netemeyer and Boles (1997) was used to assess buyer fairness. Accuracy was measured using a ten-item scale that expounded on a scale developed by Kinicki et al., (2004). Sufficiency of requirement definition was measured using a four-item scale developed by Hawkins et al., (2011). Communication bi-directionality was measured using a two-item formative scale from Mohr and Sohi (1995). Communication frequency was measured using an average of the counts of the number of communications using various media. This average count was modeled as a formative variable per Mohr and Sohi (1995).

## Pretest

In order to ensure that the constructs were valid in content and the survey items sufficiently clear, the survey instrument was reviewed by several academicians and contracting practitioners. Academicians included those from the School of Business and Public Policy at the Naval Postgraduate School.

These experts from industry and academia were asked to review the survey instrument. As recommended by Dillman (2000), feedback was solicited regarding whether the survey items: (1) captured the domain of the construct (content validity), (2) were unambiguous, (3) were simple to understand, and (4) were consistently interpretable. The experts were asked whether the model was sufficiently comprehensive, that is, whether it included all of the relevant constructs. The survey was modified to reflect improvements recommended by the experts.

## Pilot Test

In an effort to ensure construct reliability and validity, the survey instrument was pilot tested using a sample of assessing officials from defense organizations. The population included 265 assessing officials, from which 75 responded. However, 34 responses were incomplete yielding 41 usable responses and a response rate of 15.5%. Data from the pilot test was used to assess construct reliability and validity prior to full-scale survey deployment (Churchill, 1979).

## Reliability & Validity

Internal consistency reliability for each latent construct was assessed using Cronbach's alpha. All constructs showed adequate reliabilities greater than 0.7 for established scales and greater than 0.6 for new scales (Hair et al., 2010). Since the sample was less than 50, exploratory factor analysis could not be used to assess construct validity. Face validity from the pretest was deemed sufficient.

An online survey was used to collect the data. Web-based surveys yield slightly higher response rates than do mail surveys, and the data exhibits no characteristic differences than that of mail surveys (Griffis et al., 2003). The survey included approximately 145 questions (items) that measured each construct and variable in the model, including demographics collected in order to facilitate an assessment of generalizability. An email invitation was sent to respondents informing them of the purpose and importance of the research. This invitation included an embedded link to the internet universal resource locator

(URL) to facilitate convenient access to the survey. One follow-up message served as a reminder to prospective respondents.

In order to maximize the response rate, survey deployment and data collection utilized Dillman's (2000) "Tailored Design Method" for internet surveys. Dillman's method entails establishing trust with the respondent, increasing the rewards for completing the survey, and mitigating the costs of completing the survey. To establish trust in the current design, sponsorship by a legitimate authority (Dillman, 2000) was obtained as required by military department policy. The invitation identified that the research is for the purpose of grant-sponsored research, and that WKU's Institutional Review Board would maintain oversight of the research.

In order to provide rewards, the researcher showed positive regard to the respondent (Dillman, 2000). In the email invitation, respondents were referred to as valued experts whose input is critical to the research. Additionally, the invitation showed a support of group values (Dillman, 2000). The researcher was identified as a cohort in federal contracting. The respondents were offered a report of the results of the research. Finally, respondents were offered an opportunity to be entered into a raffle for an iPad Mini as appreciation for their support.

In order to reduce the perceived costs of completing the survey, the survey questions were kept relatively simple; the time required to seek information was minimized. With the exception of a few demographic questions such as gender, personal information was not requested, and responses were anonymous (Dillman, 2000).

## Full Sample

The personnel with the requisite knowledge of contractor performance were those who served as CPARS assessing officials. In pursuit of a sample size sufficient to test the model, the survey was presented to 2,247 assessing officials in defense organizations. From those invited, 148 responses were received. However, 58 of those responses were incomplete resulting in 90 usable responses. The records from the pilot study were then added to the data set. This combined sample of 131 respondents out of 2,512 resulted in a response rate of 5.2%.

### Demographics

The average dollar value of the contracts was $164.7 million (std. dev. $971.8M; range: $62K-$10B). The respondents' average years of experience

assessing contractor performance was 14.75 (std. dev. 9.5). Demographics characterizing the respondents and the contracts for which they responded are found in the ensuing tables. The sample was respectably educated. Assessing officials, although were mostly program managers, represented a variety of job functions. Respondent ages were evenly distributed across ten-year groups. Most respondents were male (72%), which is somewhat skewed compared to total U.S. government employment (57%) (Office of Personnel Management, 2014). The sample is heavily influenced by services versus construction and goods. Professional services dominate the service category. Most transactions were completed, and large and small businesses are evenly represented. All major contract types are represented; however, most are firm-fixed price and cost reimbursement. Thirty percent of contracts contained incentives (award fee, incentive fee, award term, performance-based payments, and/or liquidated damages). Tables 2-11 further describe the sample and provide insight to the extent of generalizability of the results.

Table 2. Highest Education Attained

| Degree Type | Frequency |
|---|---|
| High School | 12 |
| Associates | 8 |
| Bachelors | 31 |
| Masters | 74 |
| Doctorate | 2 |

Table 3. Assessing Official Career Field

| Group | Frequency |
|---|---|
| Quality Assurance | 3 |
| Program Management | 50 |
| Contracting | 18 |
| Engineering | 26 |
| Logistics | 12 |
| Other | 19 |

Table 4. Performance Assessing Experience

| Years | Frequency |
|---|---|
| 0 - 9 | 39 |
| 10 - 19 | 47 |
| 20 - 29 | 28 |
| 30 - 39 | 11 |
| 40 - 49 | 2 |

Table 5. Gender

| Type | Frequency | Percentage |
|---|---|---|
| Male | 91 | 71.7 |
| Female | 36 | 28.3 |

Table 6. Purchase Type

| Type | Frequency |
|---|---|
| Services | 93 |
| Construction | 4 |
| Supplies/Commodities/Spares | 17 |
| Weapon System | 1 |
| Other | 13 |

Table 7. Competition

| Type | Frequency | Percentage |
|---|---|---|
| Competed | 90 | 70.3 |
| Not Competed | 38 | 29.7 |

Table 8. Business Size

| Type | Frequency | Percentage |
|---|---|---|
| Small Business | 63 | 49.2 |
| Large Business | 65 | 50.8 |

Table 9. Type Of Contract

| Type | Frequency |
|---|---|
| Firm-Fixed Price | 77 |
| Cost Reimbursement | 38 |
| Time and Materials | 3 |
| Labor-Hour | 1 |
| Hybrid | 11 |
| Other (e.g., Basic Ordering Agreement) | 1 |

Table 10. Contract Value

| Type | Frequency |
|---|---|
| < $1 Million | 13 |
| $1 - 4.99 Million | 32 |
| $5 - 24.99 Million | 42 |
| $25 - 49.99 Million | 11 |
| $50 - 99.99 Million | 7 |
| $100 - 499.99 Million | 9 |
| $500 – 999.99 Million | 2 |
| $1 – 4.99 Billion | 2 |
| > $5 Billion | 1 |

Table 11. Product Service Code/Federal Supply Class

| PSC/FSC | Frequency |
|---|---|
| A - Research and Development (R&D) | 11 |
| B - Special Studies and Analyses | 2 |
| C - Architect and Engineering Services | 4 |
| D - Automatic Data Processing and Telecommunication | 11 |
| F - Natural Resources Management | 1 |
| J - Maintenance, Repair, and Rebuilding of Equipment | 12 |
| K - Modification of Equipment | 1 |
| L - Technical Representative | 2 |
| M - Operation of Government Owned Facilities | 1 |
| Q - Medical Services | 3 |
| R - Professional, Administrative and Management Support | 34 |
| S - Utilities and Housekeeping Services | 3 |
| U - Education and Training | 1 |
| V - Transportation, Travel and Relocation | 4 |
| W - Lease or Rental of Equipment | 1 |
| Y - Construction of Structures and Facilities | 1 |
| Z - Maintenance, Repair or Alteration of Real Property | 2 |
| 10 - Weapons | 3 |
| 12 - Fire Control Equipment | 2 |
| 14 - Guided Missiles | 1 |
| 15 - Aircraft and Airframe Structural Components | 6 |

Table 11. Product Service Code/Federal Supply Class (continued)

| PSC/FSC | Frequency |
|---|---|
| 16 - Aircraft Components and Accessories | 4 |
| 17 - Aircraft Launching, Landing, and Ground Handling Equipment | 1 |
| 18 - Space Vehicles | 1 |
| 19 - Ships, Small Craft, Pontoons, and Floating Docks | 2 |
| 28 - Engines, Turbines, and Components | 1 |
| 35 - Service and Trade Equipment | 1 |
| 39 - Materials Handling Equipment | 1 |
| 58 - Communication, Detection and Coherent Radiation Equipment | 2 |
| 59 - Electrical and Electronic Equipment Components | 1 |
| 69 - Training Aids and Devices | 1 |
| 70 - Automated Data Processing Equipment (Including Firmware), Software, Supplies, and Support Equipment | 4 |
| 71 - Furniture | 1 |
| 72 - Household and Commercial Furnishings and Appliances | 1 |
| 91 - Fuels, Lubricants, Oils, and Waxes | 1 |
| 99 - Miscellaneous | 3 |

## Measure Evaluation

### Normality

Tests for skewness and kurtosis revealed that most item z-scores for skewness were greater than an absolute value of three, suggesting that the data for most items was skewed (Kline, 1997). Skewness was visually apparent from the histogram of each item. Conversely, only items measuring fairness and one item measuring rating dissonance (D4) showed an absolute value of kurtosis z-scores greater than ten (Kline, 1997). PLS SEM is a non-parametric method; it does not require that data be normally distributed. "PLS-SEM is suitable for applications where strong assumptions cannot be fully met and is often referred to as a distribution-free 'soft modeling approach'" (Hair et al., 2012, p. 416). As a test, each scale item of the ten latent constructs that violated normality was transformed to fall within the thresholds stated above. A variety of transformations were used including squared, cubed, $4^{th}$ power, $5^{th}$ power, Log10, and inverse; however, the same transformation was used for all items of any single construct in order to keep each scale consistent. The model below was re-run with more normalized items. The path coefficient effects and statistical significances were nearly identical. Data transformations made no appreciable change in the model. Thus, all ensuing analyses are based on non-transformed data.

### Bias

A major concern in cross-sectional survey research is response bias, particularly coverage bias, selection bias, non-response bias (Blair and Zinkhan, 2006), and socially desirable responding (SDR). Coverage bias occurs when, due to research methods, a particular group is excluded from the population (Blair and Zinkhan, 2006). This research design excluded for-profit sector buyers; thus, results will need to be examined carefully prior to generalizing to this context. Otherwise, considering the breadth of demographic representation shown above, coverage bias is not a concern.

Non-response bias occurs when a particular group(s) fails to respond to the survey. Non-response bias was evaluated by comparing responses from early and late respondents. The rationale for this approach is that late respondents sufficiently resemble non-respondents (Armstrong and Overton, 1977). A chi-square test showed no difference across a key demographic, gender. Independent samples t-tests explored any differences in 15 constructs measured by continuous measures. Only one difference was found (in role

overload, which seems logical; busy people might have procrastinated). These results suggests that the sample was not affected by a non-response bias.

SDR is "the tendency to give answers that make the respondent look good" (Paulhus, 1991, p. 17). This natural tendency may obfuscate the truth; thus, SDR can seriously jeopardize the validity of survey research (Randall and Fernandes, 1991; Nunnally, 1978). "SD[R] can act as (1) an unmeasured variable that produces spurious correlations between study variables, (2) a suppressor variable that hides relationships, or (3) a moderator variable that conditions the relationship between two other variables" (Ganster et al.,1983, p. 321). Some tools are available to the researcher to control the influence of SDR (Paulhus, 1991; Randall and Fernandes, 1991). This research included a demand reduction technique (anonymity) to reduce the respondent's motivation to respond in a socially acceptable way. The research design collected data anonymously. This is consistent with other similar research of situations encountered by procurement professionals making procurement-related decisions (Landeros and Plank, 1996).

Since cross-sectional survey-based data entails multiple variables measured from a single source, common method bias must be of concern. Harmon's one-factor test showed that when latent-indicator items were forced onto a single factor in exploratory factor analysis (EFA), the 58 items accounted for 33.13% of the variance in the common method factor, which is significantly less than 50% recommended (Podsakoff and Organ, 1986). Therefore, common method bias is not great enough to affect the results.

## Reliability and Validity

To assess reliability and validity, first, a measurement model of constructs measured by reflective indicators was run using partial least squares (PLS) structural equation modeling (SEM). Similar to the pilot test, the reliability of latent constructs was assessed using composite reliability, a measure of internal consistency reliability. The composite reliability of each construct (Table 14) was compared to the generally accepted standard of 0.7 for established scales and 0.6 for new scales (Nunnally, 1978). Each construct exceeded the 0.7 threshold.

Reliability is a necessary, but insufficient, condition for validity (Kerlinger and Lee, 2000). Another aspect of validity that must be satisfied is to ensure that what is actually measured corresponds with what was intended to be measured. This aspect of validity addresses the accuracy of the measures.It was assessed via construct, convergent, and discriminant validity. Specifically, construct validity was assessed using principle components EFA with a Varimax rotation. All predictor constructs were run together in an EFA. Individual items were

assessed for sufficient correlation with the factor (factor loading), greater than 0.6, while simultaneously ensuring cross-loadings were less than 0.4. Items were iteratively trimmed until these thresholds were met. In a confirmatory manner (rather than exploratory), the items were forced on to the hypothesized number of factors. However, leverage attitude clearly split into two separate factors.

Convergent validity was established by examining average variance extracted (AVE). The AVE for each construct far exceeded the 0.5 threshold (Fornell and Larcker, 1981). Convergent validity was further assessed by examining the completely standardized factor loadings (Table 12). All loadings are statistically significant and all but one (SPE5) exceed the recommended .50 level (Hair et al., 2010). SPE5 was retained for its theoretical value in expounding on the meaning of SPE by incorporating the concept of supplier motivation. Discriminant validity was established by examining the squared correlation between each pair of constructs compared to the AVE for each associated construct (Fornell and Larcker, 1981). In each case, the AVE is significantly greater than the squared correlations (Table 13). Discriminant validity was also examined using the heterotrait-monotrait (HTMT) ratio of correlations (Henseler et al., 2015) that compares the ratio of the within-construct correlations to the between-construct correlations. All HTMT ratios were less than the recommended 0.85. Overall, the constructs were deemed to be of sufficient reliability and construct validity. Table 14 presents the means, standard deviations, scale reliabilities, and correlations for these constructs.

Table 12. Factor Loadings

| | Accuracy | Comm. Formality | Fairness | Fear of Dispute | Leverage Attitude | SPE Efficacy | Perceived Usefulness | Rating Dissonance | Rating Justification | Relationship Quality | Role Overload | Sufficiency of Requirement Definition | Surveillance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A10 | 0.904 | | | | | | | | | | | | |
| A6 | 0.805 | | | | | | | | | | | | |
| A7 | 0.852 | | | | | | | | | | | | |
| A9 | 0.881 | | | | | | | | | | | | |
| BF1 | | | 0.927 | | | | | | | | | | |
| BF2 | | | 0.952 | | | | | | | | | | |
| BF3 | | | 0.907 | | | | | | | | | | |
| BF4 | | | 0.934 | | | | | | | | | | |
| BF5 | | | 0.872 | | | | | | | | | | |
| CForm2 | | 0.954 | | | | | | | | | | | |
| CForm3 | | 0.962 | | | | | | | | | | | |
| CForm5 | | 0.896 | | | | | | | | | | | |
| FD1 | | | | 0.794 | | | | | | | | | |
| FD2 | | | | 0.816 | | | | | | | | | |
| FD3 | | | | 0.708 | | | | | | | | | |
| FD4 | | | | 0.879 | | | | | | | | | |
| L1 | | | | | 0.902 | | | | | | | | |
| L2 | | | | | 0.89 | | | | | | | | |
| D1 | | | | | | | 0.93 | | | | | | |
| D3 | | | | | | | 0.94 | | | | | | |
| D4 | | | | | | | 0.871 | | | | | | |
| SPE1 | | | | | | | | | | | 0.859 | | |
| SPE2 | | | | | | | | | | | 0.902 | | |
| SPE3 | | | | | | | | | | | 0.878 | | |
| SPE4 | | | | | | | | | | | 0.936 | | |

Table 12. Factor Loadings (continued)

| | Accuracy | Comm. Formality | Fairness | Fear of Dispute | Leverage Attitude | SPE Efficacy | Perceived Usefulness | Rating Dissonance | Rating Justification | Relationship Quality | Role Overload | Sufficiency of Requirement Definition | Surveillance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SPE5 | | | | | | | | | | | 0.414 | | |
| SPE6 | | | | | | | | | | | 0.916 | | |
| PU1 | | | | | | 0.908 | | | | | | | |
| PU10 | | | | | | 0.818 | | | | | | | |
| PU2 | | | | | | 0.94 | | | | | | | |
| PU3 | | | | | | 0.829 | | | | | | | |
| PU4 | | | | | | 0.936 | | | | | | | |
| PU9 | | | | | | 0.819 | | | | | | | |
| RD1 | | | | | | | | | | | | 0.918 | |
| RD2 | | | | | | | | | | | | 0.925 | |
| RD3 | | | | | | | | | | | | 0.902 | |
| RD4 | | | | | | | | | | | | 0.923 | |
| RD5 | | | | | | | | | | | | 0.903 | |
| RJ2 | | | | | | | | 0.92 | | | | | |
| RJ3 | | | | | | | | 0.898 | | | | | |
| RJ5 | | | | | | | | 0.908 | | | | | |
| RO1 | | | | | | | | | | 0.829 | | | |
| RO2 | | | | | | | | | | 0.893 | | | |
| RO3 | | | | | | | | | | 0.845 | | | |
| RO4 | | | | | | | | | | 0.906 | | | |
| RQS2 | | | | | | | | | 0.929 | | | | |
| RQS3 | | | | | | | | | 0.922 | | | | |
| RQS5 | | | | | | | | | 0.711 | | | | |
| RQT1 | | | | | | | | | 0.916 | | | | |
| RQT2 | | | | | | | | | 0.946 | | | | |

Table 12. Factor Loadings (continued)

| | Accuracy | Comm. Formality | Fairness | Fear of Dispute | Leverage Attitude | SPE Efficacy | Perceived Usefulness | Rating Dissonance | Rating Justification | Relationship Quality | Role Overload | Sufficiency of Requirement Definition | Surveillance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RQT3 | | | | | | | | | 0.951 | | | | |
| RQT4 | | | | | | | | | 0.922 | | | | |
| RQT5 | | | | | | | | | 0.871 | | | | |
| RQT6 | | | | | | | | | 0.934 | | | | |
| S1 | | | | | | | | | | | | | 0.933 |
| S2 | | | | | | | | | | | | | 0.942 |
| S3 | | | | | | | | | | | | | 0.959 |
| S4 | | | | | | | | | | | | | 0.933 |

## Table 13. Discriminant Validity

| | Accuracy | Comm. Formality | Buyer Fairness | Fear of Dispute | Leverage Attitude | SPE Efficacy | Perceived Usefulness | Rating Dissonance | Rating Justification | Relationship Quality | Role Overload | Sufficiency of Requirement Definition | Surveillance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy | 0.74 | | | | | | | | | | | | |
| Comm. Formality | 0.18 | 0.88 | | | | | | | | | | | |
| Fairness | 0.40 | 0.12 | 0.84 | | | | | | | | | | |
| Fear of Dispute | 0.17 | 0.11 | 0.18 | 0.64 | | | | | | | | | |
| Leverage Attitude | 0.06 | 0.10 | 0.00 | 0.11 | 0.80 | | | | | | | | |
| SPE Efficacy | 0.05 | 0.21 | 0.05 | 0.03 | 0.01 | 0.77 | | | | | | | |
| Perceived Usefulness | 0.08 | 0.00 | 0.11 | 0.23 | 0.01 | 0.01 | 0.84 | | | | | | |
| Rating Dissonance | 0.32 | 0.16 | 0.29 | 0.08 | 0.01 | 0.10 | 0.04 | 0.83 | | | | | |
| Rating Justification | 0.17 | 0.17 | 0.19 | 0.27 | 0.03 | 0.03 | 0.08 | 0.16 | 0.82 | | | | |
| Relationship Quality | 0.07 | 0.14 | 0.03 | 0.15 | 0.05 | 0.04 | 0.10 | 0.08 | 0.06 | 0.76 | | | |
| Role Overload | 0.20 | 0.23 | 0.21 | 0.11 | 0.01 | 0.15 | 0.08 | 0.27 | 0.27 | 0.09 | 0.70 | | |
| Sufficiency of Requirement Definition | 0.43 | 0.14 | 0.32 | 0.14 | 0.02 | 0.06 | 0.09 | 0.31 | 0.18 | 0.13 | 0.31 | 0.84 | |
| Surveillance | 0.09 | 0.19 | 0.22 | 0.06 | 0.01 | 0.09 | 0.02 | 0.10 | 0.12 | 0.01 | 0.16 | 0.05 | 0.89 |

AVE is shown on the diagonal.

Table 14. Construct Means, Standard Deviations, Scale Reliabilities[ab] and Correlations

| Construct | Mean | SD | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Rating Justification | 5.72 | 1.3 | (.80) | | | | | | | | | | | | | | | | |
| 2. Surveillance | 5.36 | 1.52 | .32** | (.97) | | | | | | | | | | | | | | | |
| 3. Communication Formality | 5.0 | 1.79 | .40** | .43** | (.96) | | | | | | | | | | | | | | |
| 4. Role Overload | 3.20 | 2.02 | -.26** | -.09 | -.36** | (.92) | | | | | | | | | | | | | |
| 5. Perceived Usefulness | 3.71 | 1.93 | .29** | .29** | .45** | -.17* | (.95) | | | | | | | | | | | | |
| 6. Relationship Quality | 5.69 | 1.51 | .40** | .35** | .40** | -.23** | .18* | (.97) | | | | | | | | | | | |
| 7. Buyer Fairness | 6.41 | 0.95 | .54** | .46** | .34** | -.17 | .22* | .43** | (.97) | | | | | | | | | | |
| 8. Leverage Attitude[b] | 2.53 | 1.97 | -.09 | -.08 | -.32** | .19* | -.08 | -.16 | -.07 | (.90) | | | | | | | | | |
| 9. Fear of Supplier Dispute | 2.15 | 1.69 | -.28** | -.24** | -.32** | .35** | -.16 | -.51** | -.42** | .32** | (.88) | | | | | | | | |
| 10. Perceived Accuracy | 5.57 | 2.36 | .56** | .30** | .42** | -.25** | .21* | .41** | .62** | -.24** | -.40** | (.92) | | | | | | | |

Table 14. Construct Means, Standard Deviations, Scale Reliabilities[ab] and Correlations (continued)

| Construct | Mean | SD | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11. Rating Dissonance | 2.21 | 1.66 | -.19* | -.15 | -.07 | .27** | .09 | -.27** | -.33** | .07 | .47** | -.29** | (.94) | | | | | | |
| 12. Sufficiency of Requirement Definition | 5.22 | 1.54 | .55** | .22* | .38** | -.35** | .22* | .41** | .56** | -.14 | -.37** | .65** | -.30** | (.97) | | | | | |
| 13. SPE Efficacy | 5.47 | 1.29 | .48** | .40** | .49** | -.22* | .44** | .50** | .40** | -.07 | -.29** | .38** | -.21* | .50** | (.93) | | | | |
| 14. Communication Bi-directionality[b] | 5.15 | 1.65 | .38** | .42** | .44** | -.05 | .25** | .30** | .25** | -.11 | -.08 | .31** | .04 | .22* | .19* | (.72) | | | |
| 15. Communication Frequency[c] | 3.92 | 1.08 | .29** | .27** | .22* | .01 | .22* | .09 | .10 | .05 | .11 | .13 | .12 | .10 | .20* | .52** | - | | |
| 16. Turnover[cd] | .480 | .502 | -.08 | -.11 | -.06 | .17 | -.07 | -.20 | -.07 | -.04 | .12 | .00 | .05 | -.12 | .11 | -.10 | -.15 | - | |
| 17. Hours[c] | 18.1 | 21.7 | -.06 | -.11 | -.03 | -.05 | .01 | -.19* | -.29** | .07 | .13 | -.35** | .22* | -.27** | -.14 | .03 | .11 | -.06 | - |

**Significant at the 0.01 level (2-tailed). *Significant at the 0.05 level (2-tailed). [a]Composite Reliabilities are presented on the diagonal. [b]Spearman-Brown split-half reliability for 2-item scale. [c]Single-item scale. [d]Binary variable.

THIS PAGE LEFT INTENTIONALLY BLANK

## Results - Quantitative

The model was tested using partial least squares (PLS) structural equation modeling (SEM). PLS SEM, versus covariance-based SEM, is the valid modeling approach when the model includes formative scales (Hair et al., 2014). PLS SEM also accommodates complex models with a large number of variables, can model non-normally distributed data, and does not pose problems with convergence often found in covariance-based SEM. It is also more appropriate for small sample sizes and models with binary predictor variables – both of which characterize this research (Hair et al., 2014). Given the small sample size, the research design was examined to ensure sufficient power to detect hypothesized effects. Seeking power of 80 percent, assuming a significance level of .05 and a minimum R-squared of .25, and ten times the maximum number of predictor parameters per construct to be estimated (in this case, eight), the minimum sample size was 80 (Hair et al., 2014). Thus, the sample of 133 cases was deemed sufficient to apply PLS SEM to test the model. Results are shown in Table 15.

In assessing the PLS SEM model, first multi-collinearity must be checked. Since no variance inflation factors exceeded the threshold of 5, multi-collinearity posed no concern.

Table 15. PLS Results of Estimated Path Coefficients and Effects

| | Standardized Path Coefficients | | |
|---|---|---|---|
| | Hypothesis | Standardized Path Coefficient | Hypothesis (not) Supported |
| *Direct Effects* | | | |
| Rating Dissonance→SPE Efficacy | H1 | -.15* | S |
| Rating Dissonance→Hours | H2 | .13 | NS |
| Perceived Accuracy→Hours | H3 | -.29** | S |
| Perceived Accuracy →Fear of Supplier Dispute | H4 | -.23** | S |
| Fear of Supplier Dispute→Rating Dissonance | H5 | .38*** | S |
| Sufficiency of Requirement Definition→Rating Dissonance | H6 | -.07 | NS |
| Sufficiency of Requirement Definition→Perceived Accuracy | H7 | .41*** | S |
| Turnover→Perceived Accuracy | H8 | .07 | NS |
| Perceived Accuracy→Rating Dissonance | H9 | -.05 | NS |
| Perceived Accuracy→SPE Efficacy | H10 | .17 | NS |

| | | | |
|---|---|---|---|
| Communication Frequencey (binary)→ Perceived Accuracy | H11 | -.03 | NS |
| Communication Bi-directionality(binary)→ Perceived Accuracy | H12 | .12 | PS†† |
| Communication Formality(binary)→ Perceived Accuracy | H13 | .08 | PS† |
| Role Overload→ Rating Justification | H14 | -.09 | PS† |
| Role Overload→Rating Dissonance | H15 | .13 | PS†† |
| Perceived Usefulness→Rating Justification | H16 | .17*** | S |
| Perceived Accuracy→Rating Justification | H17 | .51*** | S |
| Rating Justification →SPE Efficacy | H18 | .40*** | S |
| Perceived Accuracy→Perceived Usefulness | H19 | .22*** | S |
| Relationship Quality→Fear of Supplier Dispute | H20 | -.43*** | S |
| Communication Frequencey→ Relationship Quality | H21 | -.07 | NS |
| Communication Bi-directionality→Relationship Quality | H22 | .29** | S |
| Communication Formality→Relationship Quality | H23 | .31*** | S |
| Turnover→Relationship Quality | H24 | -.16** | S |
| Surveillance→Perceived Accuracy | H25 | -.03 | NS |
| Fairness→Perceived Accuracy | H26 | .36*** | S |
| Leverage Attitude→Perceived Accuracy | H27 | -.14** | S |
| | *Variance Explained (adjusted $R^2$)* | $Q^2$ | |
| SPE Efficacy | 31% | .15 | |
| Rating Dissonance | 23% | .20 | |
| Perceived Accuracy | 56% | .41 | |
| Perceived Usefulness | 4% | .03 | |
| Rating Justification | 35% | .29 | |
| Fear of Supplier Dispute | 31% | .18 | |
| Relationship Quality | 25% | .21 | |
| Hours | 11% | .08 | |

*$p$< .10; **$p$< .05; ***$p$< .01

†Partially supported via ANOVA, explained below; ††Partially supported via regression, explained below.

Next, effect sizes were evaluated. $Q^2$ is a measure of the model's out-of-sample predictive relevance. As seen in Table 15, all $Q^2$ values are greater than zero, indicating predictive relevance of the exogenous construct predicting the endogenous (i.e., dependent) constructs in the model. The effect size standards for $Q^2$ are the same as those for $f^2$. The $f^2$ statistics indicate the effects as follows: .02 values are small, .15 values are medium, and .35 are large (Hair et al., 2014). The $f^2$ statistics (Tables 16-19) indicate how much a particular relationship contributes to the percentage of variance explained in the dependent variable (i.e., the coefficient of multiple determination, adjusted $R^2$).

Overall, support was found for 15 of the 27 hypotheses. Of the significant relationships, one effect size was large (perceived accuracy→rating justification), four were medium (rating justification→SPE efficacy; relationship quality→fear of supplier dispute; sufficiency of requirement definition→perceived accuracy; and fairness→perceived accuracy), and the remaining ten relationships were small. The effect sizes indicate the most impactful paths (i.e., chain of relationships) to SPE efficacy. It appears that the most impactful constructs explaining SPE efficacy (i.e., its chain of effects) involve rating justification, perceived accuracy, sufficiency of requirement definition, and fairness. While others are also significant, their effects are relatively smaller.

Similar to ordinary least square regression, PLS SEM path coefficients represent the estimated change in the endogenous construct per unit of change in a predictor construct. Examining the standardized path coefficients, the effect of rating justification on SPE efficacy is more than twice that of rating dissonance. Examining the effects on rating justification, perceived accuracy has a far greater effect than does perceived usefulness. Further, perceived accuracy also affects perceived usefulness. Looking further back in the model, examining the effects on perceived accuracy, the sufficiency of the requirement definition has the greatest impact, followed closely by fairness. Leverage attitude also affects perceived accuracy, but not nearly as strongly. Fear of supplier dispute is affected by relationship quality. Rating dissonance is affected only by fear of supplier dispute.

The effect of accuracy on SPE efficacy appears to be fully mediated by rating justification. In addition, although the perceived usefulness at least partially mediates the relationship between perceived accuracy and rating justification, the total effect of perceived accuracy on rating justification dominates. Similarly, perceived accuracy does not affect rating dissonance directly, but does through a mediated relationship with fear of supplier dispute.

Table 16. Effect Sizes

| | SPE Efficacy | | Rating Dissonance | |
|---|---|---|---|---|
| | Path Coefficient | $f^2$ | Path Coefficient | $f^2$ |
| Rating Dissonance | **-.15** | .03 | | |
| Perceived Accuracy | .17 | .03 | -.05 | .00 |
| Rating Justification | **.40** | .16 | | |
| Role Overload | | | .13 | .02 |
| Fear of Supplier Dispute | | | **.38** | .14 |
| Sufficiency of Requirement Definition | | | -.07 | .00 |

Table 17. Effect Sizes

| | Rating Justification | | Perceived Usefulness | |
|---|---|---|---|---|
| | Path Coefficient | $f^2$ | Path Coefficient | $f^2$ |
| Perceived Usefulness | **.17** | .04 | | |
| Perceived Accuracy | **.51** | .36 | **.22** | .05 |
| Role Overload | -.09 | .01 | | |

Table 18. Effect Sizes

| | Hours | | Relationship Quality | |
|---|---|---|---|---|
| | Path Coefficient | $f^2$ | Path Coefficient | $f^2$ |
| Rating Dissonance | .13 | .02 | | |
| Perceived Accuracy | **-.29** | .09 | | |
| Communication Frequency | | | -.07 | .01 |
| Communication Bi-directionality | | | **.29** | .09 |
| Communication Formality | | | **.31** | .11 |
| Turnover | | | **-.16** | .03 |

Table 19. Effect Sizes

| | Perceived Accuracy | | Fear of Supplier Dispute | |
|---|---|---|---|---|
| | Path Coefficient | $f^2$ | Path Coefficient | $f^2$ |
| Relationship Quality | | | **-.43** | .22 |
| Perceived Accuracy | | | **-.23** | .07 |
| Sufficiency of Requirement Definition | **.41** | .26 | | |
| Turnover | .07 | .01 | | |
| Fairness | **.36** | .18 | | |
| Leverage Attitude | **-.14** | .04 | | |
| Communication Frequency | -.03 | .00 | | |
| Communication Bi-directionality | .12 | .03 | | |
| Communication Formality | .08 | .01 | | |
| Surveillance | -.03 | .00 | | |

## Post Hoc Analyses

Table 14 displays construct means and standard deviations. Overall, respondents in the sample reported SPEs as *somewhat* accurate (mean 5.57), and this construct varied more than any other (SD 2.36). Of the 131 respondents, 91 agreed or strongly agreed that SPEs were effective, whereas 40 (30.5%) expressed doubt. The sample also exhibited much variance in role overload (SD 2.02). Leverage attitude was somewhat low across the sample (mean 2.53), but it also varied highly (SD 1.97), as did perceived usefulness (SD 1.93). Respondents overall were slightly less than neutral concerning the utility of the CPAR process and supporting information technology system, and this sentiment varied within the sample (mean 3.71, SD 1.93). Also, respondents overall did not exhibit a fear of supplier dispute; however, this varied (SD 1.69). Respondents believed that their evaluations were highly fair (mean 6.41) and this varied little (SD 0.95). Respondents also believed their ratings were *somewhat* justified (mean 5.72, SD 1.3). Finally, in general, respondents believe SPEs are *somewhat* effective (mean 5.47, SD 1.29). Additionally, the mean of the component of SPE efficacy that gauges the supplier's motivation to perform (4.37) was noticeably lower than the mean of the overall SPE efficacy construct. Thus, some evidence suggests that respondents are less confident that SPEs are effective in motivating supplier performance than they are that SPEs mitigate the risk of future adverse selection. Further, respondents were asked to rate the extent to which an SPE rating influenced an award decision of the most recent source selection in which they participated (and in which past performance was an evaluation criterion). Responses were neutral (mean 4.35) but varied (SD 2.21).

As an intended indicator of accuracy (but was trimmed out during EFA), respondents were asked to rate the level of inflated ratings. The average was low (1.95), but responses varied (SD 1.53). Thus, groups were created of high ($n$=22) and low ($n$=109) inflation (cut point 4 on a 7-point scale). Groups were then created of high ($n$=96) and low ($n$=35) SPE Efficacy. Crosstabulation analysis examined differences between the actual versus expected counts of responses that had, for example, high inflation and low SPE efficacy and high inflation and high SPE efficacy. The differences between actual and expected counts were significant ($\chi 2 = 4.74$, $p<.05$). There were more actual counts of high inflation and low SPE efficacy than expected, and simultaneously lower actual counts than expected counts in the high inflation and high SPE efficacy groups (Table 20) suggesting a relationship between rating inflation and SPE efficacy. Furthermore, an independent samples t-test showed that the mean value of SPE efficacy among the high inflation group was 4.40 lower (using a summated, 6-item scale) than the low inflation group ($p<.01$).

Table 20. Rating Inflation & SPE Efficacy Crosstabs

**Inflbin * SPEEffbin Crosstabulation**

| | | | SPEEffbin | | |
|---|---|---|---|---|---|
| | | | .00 | 1.00 | Total |
| Inflbin | .00 | Count | 25 | 84 | 109 |
| | | Expected Count | 29.1 | 79.9 | 109.0 |
| | | % within Inflbin | 22.9% | 77.1% | 100.0% |
| | 1.00 | Count | 10 | 12 | 22 |
| | | Expected Count | 5.9 | 16.1 | 22.0 |
| | | % within Inflbin | 45.5% | 54.5% | 100.0% |
| Total | | Count | 35 | 96 | 131 |
| | | Expected Count | 35.0 | 96.0 | 131.0 |
| | | % within Inflbin | 26.7% | 73.3% | 100.0% |

Inflated ratings appear to also relate to perceived accuracy. An independent samples t-test showed that the mean value of perceived accuracy among the high inflation group was 4.45 lower (using a summated, 4-item scale) than the low inflation group ($p$<.001). Similarly, highly inflated ratings appear to be related to rating dissonance. An independent samples t-test showed that the mean value of rating dissonance among the high inflation group was 4.21 greater (using a summated, 3-item scale) than the low inflation group ($p$<.001).

Inflated ratings are also related to fear of supplier dispute. A logistic regression model was run with a binary dependent variable, inflation (1=high; 0=low). This dependent variable was regressed on fear of supplier dispute and leverage attitude. The omnibus chi-Square test ($\chi 2 = 46.88$, $p < .001$) was significant. Additionally, the Hosmer and Lemeshow test ($\chi 2 = 8.21$, $p < .223$) showed no difference between predicted and actual classifications. The portion of explained variance was respectable as evidenced by the Cox and Snell $R^2$ of .16 and the Nagelkerke $R^2$ of .27. Practical significance was evidenced by a reasonable hit ratio of 83.2%. Fear of supplier dispute had a significant effect at the .05 level (Table 21). Since these variables were not normally distributed, they were transformed. The transformed variables were then tested (Table 22) yielding the same results. The transformed variables rendered the sizes of the Beta coefficients miniscule and uninterpretable. Reflecting back to the untransformed Beta values, the model shows that fear of supplier dispute is positively related to rating inflation ($B = 1.24$).

Table 21. Effect of Fear of Supplier Dispute on Rating Inflation

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | FearofDispute | .217 | .051 | 18.405 | 1 | .000 | 1.243 |
| | Leverage | .109 | .075 | 2.146 | 1 | .143 | 1.115 |
| | Constant | -4.538 | .766 | 35.064 | 1 | .000 | .011 |

a. Variable(s) entered on step 1: FearofDispute, Leverage.

Table 22. Effect of Fear of Supplier Dispute on Rating Inflation - Transformed

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | FearInv | -15.605 | 4.762 | 10.738 | 1 | .001 | .000 |
| | LevLog10 | 1.223 | .897 | 1.859 | 1 | .173 | 3.397 |
| | Constant | -.421 | .921 | .209 | 1 | .648 | .656 |

a. Variable(s) entered on step 1: FearInv, LevLog10.

Respondents were also asked to rate agreement with the following statement: "It is futile to report the real ratings that the contractor deserves since management will either change the ratings or make me change the ratings." While the mean level of agreement was low (2.23), responses varied (SD 1.74). Thirteen respondents rated this question as a 6 or 7 (7 = strongly agree). Twenty five respondents (19%) reported that someone on the buyer team either changed or influenced a change to the SPE for reasons shown in Figure 2. Of those, nine evaluators (36%) disagreed with the change made. Note that a respondent may have more than one reason. Thus, the 13 SPEs that had insufficient facts may be the same 13 for which a supplier's rebuttal had merit. In any event, there were 10 percent of SPEs that had trouble mustering facts to support the rating(s) that the assessing official believed the supplier deserved.

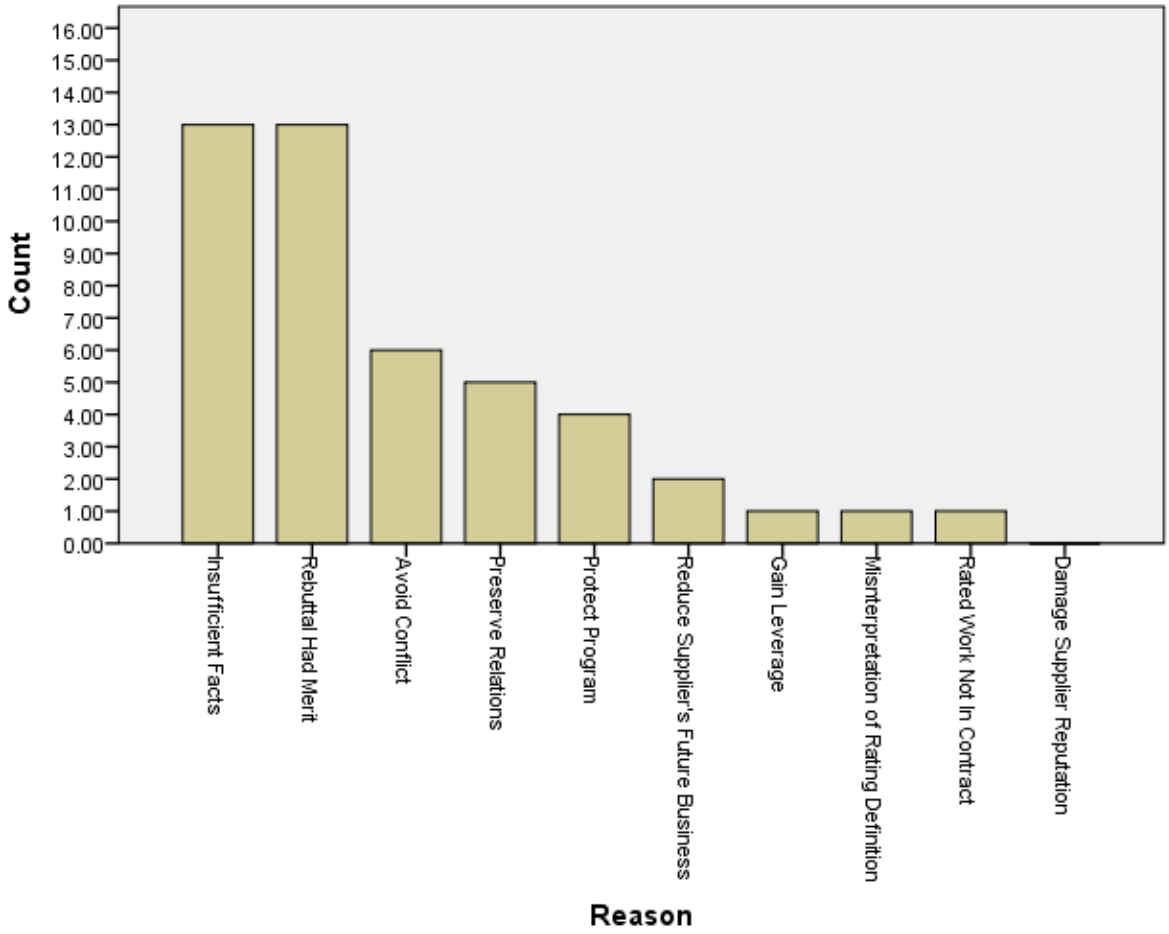


Figure 2. Reasons For Rating Changes

SPEs ratings have been reported to be incomplete – a phenomenon that was confirmed in the sample. Forty four respondents (33.5%) reported that their SPE included at least one rating category that was not complete. Groups were created of high/low levels of perceived usefulness, accuracy, role overload, and communication bi-directionality. Then, crosstabulation analyses (Tables 23-25) examined differences between the expected counts and actual counts in each combination groups (high/low levels of each variable and groups with and without incomplete ratings). Role overload was not significant ($\chi 2 = .091$, $p=.81$). However, perceived usefulness ($\chi 2 = 4.17$, $p<.05$), perceived accuracy ($\chi 2 = 5.67$, $p<.01$), and communication bi-directionality ($\chi 2 = 9.17$, $p<.01$) were each significantly different than expected suggesting that each is related to incomplete SPEs.

Table 23. Incomplete SPEs and Perceived Usefulness Crosstabs

**Incompletebin * PUbin Crosstabulation**

| | | | PUbin | | |
|---|---|---|---|---|---|
| | | | .00 | 1.00 | Total |
| Incompletebin | .00 | Count | 43 | 44 | 87 |
| | | Expected Count | 48.5 | 38.5 | 87.0 |
| | | % of Total | 32.8% | 33.6% | 66.4% |
| | 1.00 | Count | 30 | 14 | 44 |
| | | Expected Count | 24.5 | 19.5 | 44.0 |
| | | % of Total | 22.9% | 10.7% | 33.6% |
| Total | | Count | 73 | 58 | 131 |
| | | Expected Count | 73.0 | 58.0 | 131.0 |
| | | % of Total | 55.7% | 44.3% | 100.0% |

Table 24. Incomplete SPEs and Perceived Accuracy Crosstabs

**Incompletebin * Accuracybin Crosstabulation**

| | | | Accuracybin | | |
|---|---|---|---|---|---|
| | | | .00 | 1.00 | Total |
| Incompletebin | .00 | Count | 51 | 36 | 87 |
| | | Expected Count | 57.1 | 29.9 | 87.0 |
| | | % of Total | 38.9% | 27.5% | 66.4% |
| | 1.00 | Count | 35 | 9 | 44 |
| | | Expected Count | 28.9 | 15.1 | 44.0 |
| | | % of Total | 26.7% | 6.9% | 33.6% |
| Total | | Count | 86 | 45 | 131 |
| | | Expected Count | 86.0 | 45.0 | 131.0 |
| | | % of Total | 65.6% | 34.4% | 100.0% |

Table 25. Incomplete SPEs and Communication Bi-Directionality Crosstabs

**Incompletebin * CBbinary Crosstabulation**

| | | | CBbinary | | |
|---|---|---|---|---|---|
| | | | .00 | 1.00 | Total |
| Incompletebin | .00 | Count | 16 | 71 | 87 |
| | | Expected Count | 23.2 | 63.8 | 87.0 |
| | | % of Total | 12.2% | 54.2% | 66.4% |
| | 1.00 | Count | 19 | 25 | 44 |
| | | Expected Count | 11.8 | 32.2 | 44.0 |
| | | % of Total | 14.5% | 19.1% | 33.6% |
| Total | | Count | 35 | 96 | 131 |
| | | Expected Count | 35.0 | 96.0 | 131.0 |
| | | % of Total | 26.7% | 73.3% | 100.0% |

Several regression models were run to explore the effects of the type of buy (e.g., services, construction, commodities, spares, weapon systems, or capital equipment). Services were coded 1; others were coded zero. When rating justification was regressed against perceived usefulness, perceived accuracy, and type of buy (binary), type of buy was significant ($F$=22.66, standardized $B$=.16, $t$=2.13, $p$<.05). Thus, more extensive rating justifications were related to buying services. Together, these three predictor variables explained 33% of the variance in rating justification. Note that transformed values were used in this model due to non-normal data. Logically, controlling for the known effects of perceived accuracy, the type of buy (services) was also related to lower perceived usefulness ($F$=6.46, standardized $B$=-.22, $t$=-2.57, $p$<.01). Neither SPE Efficacy, rating dissonance, nor perceived accuracy were related to the type of buy (binary).

Twenty five respondents (19%) reported that the supplier disputed at least one rating and/or rating justification. Within the 25, 45 ratings were challenged. Post hoc tests explored factors contributing to supplier rebuttals to ratings and/or narrative justifications. A logistic regression model was run with a binary dependent variable, disagreement (1=disagreed; 0=did not disagree). This dependent variable was regressed on relationship quality, rating justification, perceived accuracy, and buyer fairness. The omnibus chi-Square test ($\chi 2$ = 21.61, $p$ < .001) was significant. Additionally, the Hosmer and Lemeshow test ($\chi 2$ =12.98, $p$ < .112) showed no difference between predicted and actual classifications. The portion of explained variance was respectable as evidenced by the Cox and Snell $R^2$ of .15 and the Nagelkerke $R^2$ of .24. Practical significance was evidenced by a reasonable hit ratio

of 80.2%. Two factors were found to be significant (Table 26) at the .05 level, relationship quality and perceived accuracy. Table 26 shows the untransformed variables tested. However, since these variables were not normally distributed, they were transformed. The transformed variables were then tested (Table 27) yielding the same results but with much lower probabilities of a type I error (i.e., greater statistical significance). The transformed variables rendered the sizes of the Beta coefficients miniscule and uninterpretable. Reflecting back to the untransformed Beta values, the model shows that higher accuracy of the SPE (B = -.102) decreases disagreement as does higher relationship quality (B = -.045).

Table 26. Logistic Regression, Supplier Disagreement – Untransformed

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Accuracy | -.102 | .062 | 2.695 | 1 | .101 | .903 |
| | Fairness | -.082 | .069 | 1.422 | 1 | .233 | .922 |
| | CommFormality | .008 | .060 | .017 | 1 | .896 | 1.008 |
| | CommBidirection | -.059 | .099 | .354 | 1 | .552 | .943 |
| | RatingJust | .045 | .085 | .282 | 1 | .595 | 1.046 |
| | RelQlty | -.045 | .020 | 4.986 | 1 | .026 | .956 |
| | Constant | 5.217 | 2.098 | 6.186 | 1 | .013 | 184.367 |

a. Variable(s) entered on step 1: Accuracy, Fairness, CommFormality, CommBidirection, RatingJust, RelQlty.

Table 27. Logistic Regression, Supplier Disagreement - Transformed

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | RQ4th | .000 | .000 | 7.036 | 1 | .008 | 1.000 |
| | RJcubed | .000 | .000 | .937 | 1 | .333 | 1.000 |
| | A4th | .000 | .000 | 3.727 | 1 | .054 | 1.000 |
| | BF5th | .000 | .000 | .896 | 1 | .344 | 1.000 |
| | CBSq | -.004 | .006 | .508 | 1 | .476 | .996 |
| | CFSq | .000 | .002 | .015 | 1 | .904 | 1.000 |
| | Constant | 1.081 | .779 | 1.927 | 1 | .165 | 2.947 |

a. Variable(s) entered on step 1: RQ4th, RJcubed, A4th, BF5th, CBSq, CFSq.

Overall, rating dissonance in the sample was low (mean 2.21); however, levels of dissonance varied (SD 1.66). Excluding the pilot study data (due to an ambiguity in the pilot survey), each contract utilized several supplier performance evaluators (mean 7.11, median 4, range 1-100). Appendix E shows reasons for rater dissonance. The dissonance pertained to differences of perspective of multiple evaluators, different interpretations of the definitions of ratings, confusion over the work (not) required by a contract, fear of supplier rebuttal, insufficient rating justification, inadequately-defined requirements, whether to not penalize a supplier that self-identifies and resolves a problem, and uncertainty as to the level of performance that constitutes exceeding the requirement in a way that is sufficiently beneficial to the buyer. From these sources of dissonance, different interpretations and definitions of ratings were by far the most prevalent.

Rating justification was also found to be affected by the sufficiency of rating definitions. A regression of the transformed values in the full sample supported a linear relationship ($F$=43.47, $p$<.001). Combined, accuracy, perceived usefulness, and rating definition explained 59% of the variance in rating justification. A sufficient rating definition had a significant, positive effect on rating justification (standardized $B$=.64, $p$<.001). Again, it should be noted that sufficiency of rating definition was measured using a single-item scale; thus, its reliability could not be determined and was thus not included in the PLS SEM model.

In the PLS SEM model, linear relationships between role overload and rating justification and between role overload and rating dissonance were not found. However, two groups were created – those with low and high role overload. There were 22 cases in the high overload group. The cut point was scores (1-7) greater than 5.0. Two independent samples t-tests showed differences in mean values of rating justification ($t$=-2.136, $p$<.05) and in rating dissonance ($t$=2.617, $p$<.01) between the two groups, suggesting an effect of role overload. The mean value of the summated scale of rating dissonance was 2.68 greater for the high role overload group compared to the low role overload group. The mean value of summated scale of rating justification was 1.56 lower for the high role overload group compared to the low role overload group. These differences are consistent with the directions of the hypothesized effects (H14 and H15); thus, partial support is found for these hypotheses. A further investigation of these relationships involved the use of polar extremes. The cases with middle values of role overload were removed from the sample (summated values of 9-19, four-item scale, $n$=76). Rating justification was regressed on polar role overload (binary coded as high/low) along with accuracy and perceived usefulness. Role overload was not significant. Next, rating dissonance was regressed on polar role overload along with perceived accuracy and fear of supplier dispute. The model was significant ($F$=9.09, $p$<.001),

and role overload was significant (standardized $B$=.24, $t$=2.12, $p$< .05). Together, these three variables accounted for 24% of the variance in rating dissonance.

In 48 cases (37%), prior to initiating the SPE, the buyer solicited input from the supplier about its view of what ratings or narrative justifications should be in the SPE. In 46 cases, the supplier provided input. Performance assessors overall relied somewhat on the suppliers' inputs (average rating 4.6 on a 7 point scale from "none" to "a lot" of reliance).

Assessors reported the number of hours consumed by the buying team completing the SPE. Responses ranged from 0.5 to 100 hours (mean 18.1, SD 21.7, median 8). These hours represent transaction costs of relying on suppliers to execute part of the agencies' missions.

The survey included an open text field for which respondents were invited to recommend improvements to the SPE policy. Figure 3 displays the most common issues along with frequencies. Accuracy, consistency, ambiguous definitions for performance criteria and ratings, and questionable utility were the most frequently mentioned concerns. The respondents' input is largely consistent with the hypotheses explored herein, offering further evidence of nomological validity.

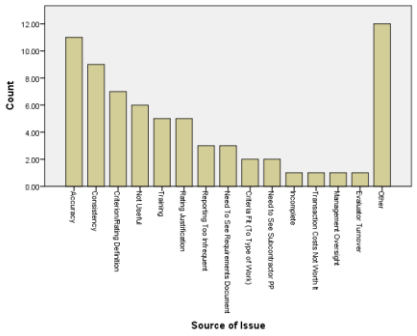


Figure 3. Assessing Officials' Issues Needing Attention

Supplier performance management was rarely augmented by information technology other than CPARS. Several assessors reported using a combination of Microsoft Excel, email, and customer surveys. Only seven respondents identified the name of a non-CPARS data system used to collect and track supplier performance information. When asked about the use of other supplier management systems, most respondents reported "none." Out-of-cycle CPARs were not common, used by only 20 assessors (15%). Figure 4 shows the different methods used to actively manage supplier performance. "Actively manage" was defined as continuous measure performance and periodically communicate the buyer's assessment of performance to the supplier to foster continuous improvements throughout the period of performance. An industry best practice, supplier scorecards, was hardly used in the sample.

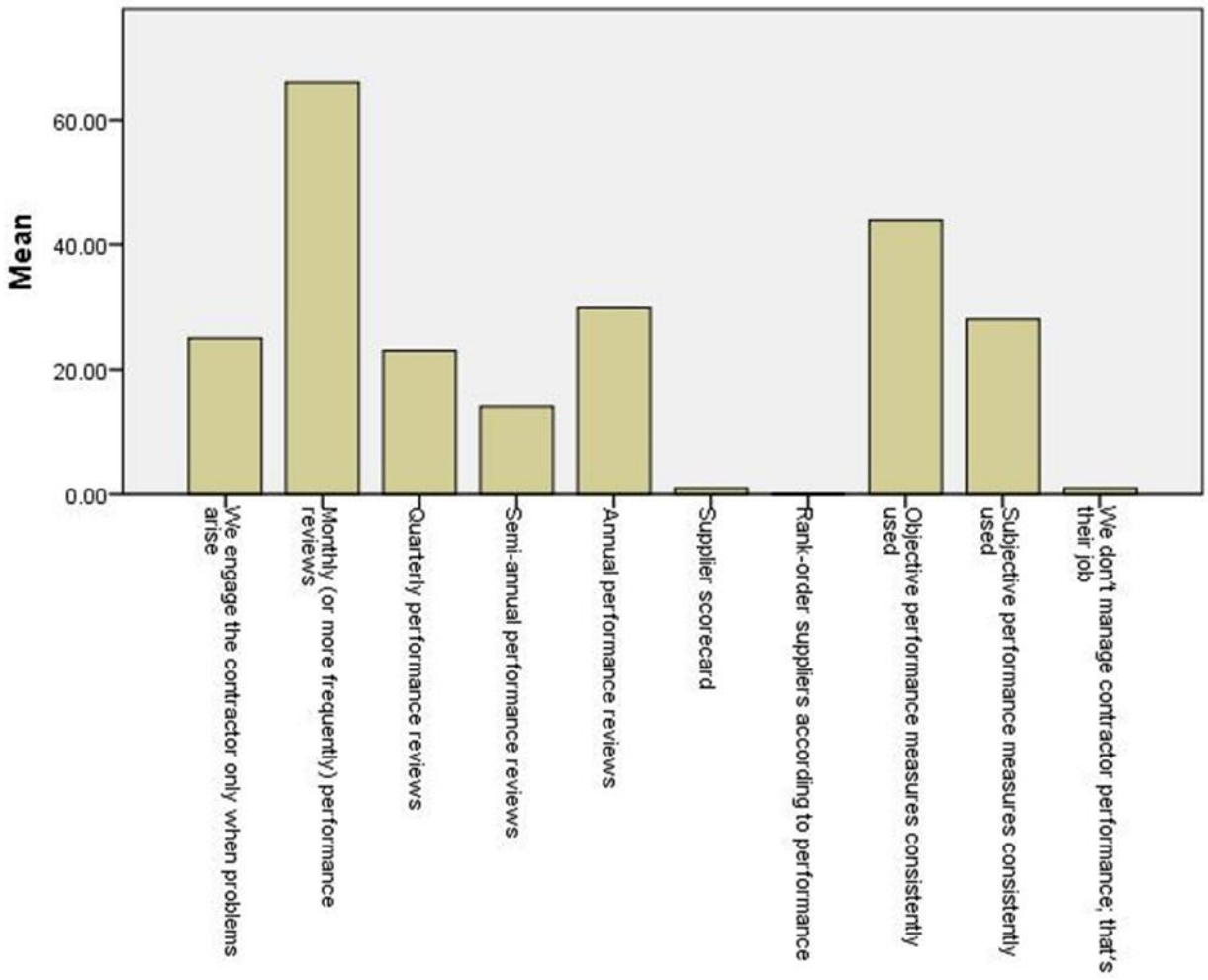


Figure 4. Performance Management Methods

Overall, respondents appeared satisfied with the suppliers' performance levels. A single-item gauge of repurchase intention reported from the CPAR showed a mean of 4.38 on a five-point scale.

Out of 131 contracts reported, 61 (46.5%) experienced turnover of at least one assessing official during the performance period. Of the 61, the average turnover per contact was 2.9 times. Total turnover ranged from 0 to 50 personnel over the life of the contract.

In the PLS SEM model, linear relationships between the three aspects of communication (bi-directionality, formality, and frequency) and perceived accuracy were not found. However, two groups were created for each aspect of communication – those with low and high communication formality, low and high communication frequency, and low and high communication bi-directionality. There were 96, 91 and 27 cases in these groups with high values, respectively. The cut

points were scores (1-7) greater than 4, 4, and 4.99, respectively. Three independent samples t-tests showed differences in mean values of perceived accuracy between the two groups of communication bi-directionality ($t$=2.917, $p$<.01) and communication formality ($t$=3.85, $p$<.01), suggesting their effects on perceived accuracy. The mean value of summated scale of perceived accuracy was 2.78 greater for the high communication bi-directionality group. Similarly, the mean value of summated scale of perceived accuracy was 3.92 greater for the high communication formality group. These differences are consistent with the direction of the hypothesized effects (H12 and H13); thus, partial support is found for these hypotheses.

A further investigation of these relationships involved the use of polar extremes. The cases with middle values of communication bi-directionality were removed from the sample (summated values of 9-12 of the two-item scale, $n$=62). Perceived accuracy was regressed on polar communication bi-directionality (binary coded as high/low) along with leverage attitude, buyer fairness, communication formality, and sufficiency of the requirement definition. The summated scale of perceived accuracy was transformed to the fourth power in order to establish normality. The model was significant ($F$=18.25, $p$<.001), and communication bi-directionality (binary) was significant (standardized $B$=.21, $t$=2.07, $p$< .05). Together, these three variables accounted for 59% of the variance in perceived accuracy. Similarly, the polar extremes of communication formality were analyzed; however, communication formality (binary) was not significant in the model.

Hypothesis 27 posited a relationship between leverage and SPE accuracy; it was supported. The pilot study survey (and the study survey) measured attitude toward using the SPE as leverage, but did not measure actual leverage employed on the contract for which data was reported. The study survey ($n$=90) included additional items to assess actual leverage employed. Attitude toward leverage and actual leverage were both measured with four item scales. In EFA, both scales' items loaded on two separate latent factors rather than on one, suggesting that the two sets of items for each scale had separate meanings. Reexamining the items (L1, L2, L3, and L4), L1 (threaten a low rating) and L2 (use SPE as bargaining leverage) seemed to correspond to ex ante threats to get the supplier to do something (i.e., *proactive leverage*). Conversely, L3 (supplier owes the buyer for an inflated rating) and L4 (leverage can be gained from an inflated SPE) seemed to correspond to a debt owed by the supplier ex post (i.e., *quid pro quo leverage*). For the attitude toward leverage construct, the average summated scale for L1 and L2 was 5.0 (possible value range 2-14), whereas the average summated scale for L3 and L4 was significantly lower at 3.3 ($t$=5.72, $p$<.001). Similarly, for the actual leverage construct, the average summated scale for AL1 and AL2 was 3.5 (possible value range 2-14), whereas the average summated scale for AL3 and AL4 was

significantly lower at 2.6 ($t$=3.45, $p$<.01). 30.5% of the 131 respondents indicated an above-neutral attitude toward leverage on at least one of the two scale items. Conversely, 7.8% of the 90 respondents in the full study indicated above-neutral actual leverage on at least one of the two scale items. Together, the data confirms a difference between an attitude toward leverage (on both proactive leverage and quid pro quo leverage) and actual leverage employed with actual leverage being significantly more rare ($t$=5.72, $p$<.001; $t$=3.41, $p$<.01, respectively). With only seven cases indicating actual leverage employed, any effects of this form of leverage could not be examined statistically. The data also confirms that evaluators treat proactive leverage and quid pro quo leverage differently with proactive leverage being more prevalent. Note that H27 was tested using the attitude toward proactive leverage. The attitude toward quid pro quo leverage was substituted into the PLS SEM model post hoc, and did not significantly affect perceived accuracy. However, non-significance is likely due to having too few cases favoring quid pro quo leverage (i.e., too little variance).

THIS PAGE LEFT INTENTIONALLY BLANK

## Qualitative Data Analysis

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 questions surrounding supplier reactions to performance evaluations met all three criteria.

## Data Collection

The interview protocol (Appendix C) was developed based on a literature review surrounding supplier performance evaluation, 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 32 and 65 minutes. Most interviews were recorded, then transcribed. Two face-to-face interviews were not recorded at the request of the informants. Transcripts averaged 13.5 pages in length. One interview occurred in-person, five occurred via telephone, and two informants provided only written testimony.

## Data Analysis

The analysis process began by identifying constructs, defining those constructs, and then positing relationships between them (Patrick Van Ecke, 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.

## Sample

The sample of informants (Table 28) was identified from awarded contracts exceeding $150 thousand and from contacts made at a trade association annual conference. Input from representatives of federal contractors who had been directly involved in the CPAR evaluation process was sought. These experts represented contractors to defense agencies across four industries. The perspectives of large and small businesses were obtained. Experience in managing customer evaluations ranged from three to 34 years, and there was a similar wide range of the number of past performance evaluations experienced (12–50). Since CPARS are used by contractors to gauge customers' perceived performance levels on current contracts and are later used in proposals pursuing new business, the

informants included both types of users (i.e., program managers and new business developers).

Table 28. Informant Demographics

| Informant | Business Size | Industry | Experience Managing Customer Evaluations (Years) | Duty Title | Supplier Performance Experience (Number of Evaluations) |
|---|---|---|---|---|---|
| 1 | L | Aerospace | 34 | Systems Engineer | Multiple |
| 2 | L | Aerospace | 7 | Program Manager | 50 |
| 3 | L | Aerospace | 14 | Program Manager | Multiple |
| 4 | S | Information Technology | 3 | COO | |
| 5 | L | Munitions | 4 | VP, Business Development | 12 |
| 6 | L | Aerospace | 30 | VP, Business Strategy | 50 |
| 7 | L | Aerospace | Multiple | Contracts Director | 24 |
| 8 | S | Shipbuilding | 30 | President | 30+ |

# Results - Qualitative

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 proposed relationships represented in the conceptual model (Figure 1).

### 1. What factors decrease the efficacy of SPEs?

This research question investigates supplier evaluation practices that either: (1) hinder a buyer's efforts to mitigate the risk of adverse selection or (2) that do not motivate suppliers to increase performance (i.e., the two purposes of SPEs). The ability to achieve the first purpose using CPAR was called into question by one informant, stating: "I have yet to see a competitive procurement that was lost because of poor CPARs." Yet another informant reported that a CPAR was *the* reason his firm did not win a contract.

One informant mentioned CPARs being reported at the basic contract level rather than at the task/delivery order level. "So far at the task order level, it does not appear that they use it." The comments highlight the variance in practices across government organizations. First, some do not report CPARs at the task order level, only at the basic contract level. This, of course, will decrease the fidelity of the information; it will be more general and less informative. More general information will make the task of future source selection teams more difficult to determine the relevancy of the past performance information. One informant commented: "You couldn't necessarily tell what [product] line they were talking about, so it was just deciphering [inaudible], you know, such a big contract having just one CPARS for it. It wasn't real clear on what data the comments were, you know, within [inaudible] they were addressing." Another informant was more critical: "We've had IDIQ level CPARS, which, honestly, I think are worthless. Because at the [inaudible] level, it's not really feedback. I think it's really for, at least from the ones we've received, the government is just checking off their box. At the IDIQ level, we've got every [inaudible], which isn't that much honestly. You know, the nuts and bolts are when you get to the task order award level. At the task order award level, that's when the CPARS really matter and you get good feedback or bad feedback."

Second, one informant reported that some organizations do not use past performance as an evaluation criterion for task order awards, thereby decreasing SPE efficacy.

One informant commented on the efficacy of CPARs: “from an overall standpoint I think the CPARS are doing what was intended for them to do.”

From the findings above emerge the following propositions:

**P1: The specificity of SPEs conducted at a parent contract level will be less than that of those conducted at a task order level.**

**P2: SPE efficacy will be positively related to SPE specificity.**

**2. How do suppliers react to inaccurate SPEs?**

As described below in the answer to research question number six, suppliers work to correct the record. They incur significant transaction costs in doing so. The buyer-supplier relationship also appears to suffer. Most suppliers reacted by rebutting the assessment in cases where they believed their rating and/or justification was not accurate. In doing so, several suppliers reported involving senior management in the reaction process. Some suppliers also reported that their company has rebutted CPARs whose ratings were negative even when they believed the CPARs ratings were warranted. When asked what would trigger a rebuttal, one informant commented: “I guess the criteria of the fairness of the feedback and it depends on the rating, right? If the rating is marginal or anything below that.” Thus, regardless of the truth, some suppliers will take measures to preserve their reputation. One supplier reacted by blaming the customer for the ultimate late delivery of the product. One informant mentioned that the relationship sours. Then, the supplier retreats to providing the customer only what is in the contract – no more. Of course, in cases in which requirements are ill defined, this supplier reaction could result in the buyer receiving less than it truly needs. One supplier reported abandoning a customer permanently. One supplier reported that its company policy is to do little or nothing regardless of what the CPAR says. It should be noted that this supplier operates largely in a non-competitive arena. Another informant reported an increase in a willingness to bill for administrative hours spent correcting the CPAR. Thus, some evidence suggests that suppliers retaliate or attempt to get even in some way. Some suppliers also reacted by preempting later CPARs processes by providing performance rating input to the assessing official prior to the CPAR due date. Several suppliers reported making a concerted effort to understand the customer’s positions, then taking added measures to better define the customer’s performance expectations. This sometimes involved refining contract documents. Then, suppliers reported taking measures to ensure those newly-refined expectations will be met in the future. No suppliers in the sample reported filing a claim to formally dispute the CPAR ratings.

One informant mentioned increasing documentation: "We will make sure that is well documented and [inaudible] that I don't get dinged on that one in the CPAR." This, absent automation, will increase transaction costs.

One informant alluded to increased transaction costs associated with duplicative past performance information submission processes and sources. The informant implied that duplication was necessary due to missing CPARs and inaccurate CPARs due to variance across government agency reporting.

> "Do they really want us to regurgitate and spit it out? And so is another thing because right now we're in a time of cost efficiencies and in trying to help the government lift them out from a bid and proposal process, and when you make me regurgitate this, it is not helping to reduce cost because I have got to somehow account for all of the B&P money that I spend on RFPs…. which would save 20 or 30 hours of not having to spend [inaudible] money on."

Based on the findings above, the following propositions are offered:

**P3: The lower the SPE efficacy, the higher the supplier's transaction costs.**

**P4: The lower the SPE efficacy, the higher the buyer's transaction costs.**

**P5: The lower the SPE accuracy, the higher the supplier's transaction costs.**

**P6: The lower the SPE accuracy, the higher the buyer's transaction costs.**

**P7: The lower the SPE accuracy, the lower the relationship quality.**

**P8: The lower the SPE accuracy, the lower the supplier's investment in transaction specific assets.**

**P9: Competition moderates the relationship between SPE accuracy and seller's transaction costs such that for suppliers selling primarily in non-competitive markets there will a lesser effect of SPE accuracy on seller's transaction costs.**

**P10: Competition moderates the relationship between SPE accuracy and buyer's transaction costs such that for suppliers selling primarily in non-competitive markets there will a lesser effect of SPE accuracy on buyer's transaction costs.**

**P11: There will be a positive relationship between a lower-than-expected SPE and future communication between the buyer and supplier.**

**P12: There will be a positive relationship between a lower-than-expected SPE and future modifications to contractual documents.**

**P13: The lower the SPE accuracy, the higher the functional conflict.**

**P14: The lower the SPE accuracy, the higher the dysfunctional conflict.**

**P15: The lower the SPE accuracy, the higher the probability of a supplier rebuttal.**

**P16: There will be a negative relationship between SPE accuracy and supplier retaliation.**

**P17: Relationship quality moderates the relationship between SPE accuracy and supplier retaliation such that in cases of low relationship quality, SPE accuracy will be negatively related to supplier retaliation (i.e., opportunism).**

**P18: A suppliers premepting of SPE evaluations (i.e., early communication) will be negatively related to the supplier's SPE rating expectation gap.**

**P19: SPE accuracy will be positively related to SPE efficacy.**

**P20: A SPE that is lower-than-expected will be positively associated with supplier reputation preservation activities (rebuttal, blame the buyer, and negotiating ratings).**

### 3. Do SPEs, in general, motivate suppliers to increase performance?

The responses to this question were mixed. One informant qualified the effect contingent on the proper use of the evaluation system as follows: "Fundamentally yes, if the system is followed and government employees have good education on the use of the system." One informant clearly backed the relationship between SPEs and performance, stating: "I think there is no doubt the secondary purpose [improved performance] is by far what I think it's most accurate or is doing the best. It is definitely motivating us. It is a metric that we are focused on. We look at it monthly across all of our contracts, we get down to—yes, they come out annually, but over our portfolio with these 50, we are looking at just as soon as the scores come out where those are…we are definitely motivated to not have any yellows or reds and we are definitely trying to be proactive in that by doing the self-assessments and not just doing them at the end of the period to help increase the score, but to try to do them at least semiannually so we have a

snapshot of where we are and if there are issues, trying to bring them up early enough that we can try to resolve them by the end of the period."

However, three informants reported no effect on their level of effort; they strive for excellence regardless. When asked whether performance evaluations motivate a change in performance levels, one informant commented: "As a major business the answer is no. We strive to do better because it is good business. The CPARs are important to us only because they are important to the customer." Another commented: "The CPAR system is written at the end of the period performance. Thus all of my performance that mattered is behind me—its past history. The only changes I can make are in the future. As a contractor we set goals every year for customer performance and most every year we exceed them. We are very conscientious to provide our customer with great service, product and performance." Another informant stated: "I'm motivated to do a good job regardless even if there was a CPAR or not." Another informant that operated in a less competitive industry was more critical, stating: "I don't think the past performance reports provide any value added for the customer."

One informant made the association between CPARs and performance effort quite clear, stating: "So we give great attention to CPAR elements that have yellow or red elements to correct those and to actively try to keep from getting those scores through better communication."

**P21 : There is a positive relationship between SPE use and supplier performance.**

### 4. How does the accuracy of SPEs affect relationship quality?

The relationship between the buyer and supplier arose throughout many of the informants' responses to many questions. When asked whether the informant suspected that the government ever uses the PP rating/evaluation as leverage, one informant answered: "Yes, especially if they don't get along with the contractor's managers." Another informant stated: "Absolutely, we have a client who we are helping now, because the government client is using this to reduce the requests for equitable adjustment." A third informant commented: "that the CPARS is used to change our position when we negotiate issues and when tough positions are brought to the forefront. "Well, you know, you only got this on your CPARS, therefore, you need to work harder so you should give us the—." So you are told to negotiate." "I think it's used as leverage every time we go to negotiate." "It's implied. It's just hideous. It's under the surface." This informant further commented: "It's a bad marriage." Another commented: "Yeah, I'd say leverage,

because I guess I don't know exactly what gets them, but they definitely use them as an opportunity to express their displeasure." One informant put it this way: "Some government program officers use the CPAR as a means to maybe get our attention or to get maybe a separate agenda, at least on the draft if not on the final version." When asked whether the government uses CPARs as leverage or a threat, one informant commented: "Oh certain—you know where we might have some thoughts that way, I don't know that it is seen as an overall trend. There are probably elements to it in some cases, but to me it is the exception. It is certainly not the rule." One informant alluded to using the CPAR as punishment: "Customers are inconsistent. They are consistent when they want to fillet you." Yet another informant commented that when compared to customer evaluations from other businesses (e.g., from prime contractors when acting as a subcontractor), the prime's SPE is more fair, more open, more forthright, less structured, and the relationships are stronger. This comment suggests that sometimes the government's CPARs may not be entirely forthright.

One informant reported that his company will no longer do business with a particular government organization due to misplaced blame on the contractor for repair delays; thus, in this case, the relationship was ruined. Another informant commented: "We as the Contractor are highly agitated. The relationship is strained." "Trust deteriorates." Another stated: "Well there is no doubt there is a strained relationship when the customer puts out a CPAR that surprises us with a negative."

**P22: SPE use increases buyer's bargaining leverage (i.e., coersive power) for the buyer.**

**P23: The greater the buyer's use of the SPE as bargaining leverage (coersive power), the lower the relationship quality.**

### 5. Why are SPEs often inaccurate?

Many factors were identified to affect the accuracy of SPEs. One informant identified biases commonly mentioned in employee performance appraisal literature – recency and an emphasis on the negative. "The natural inclination is that even though it to be over the 12 month period, they think about what has happened to them recently and they tend to think about the bad things more than the good things." Another informant reinforced this position, stating: "Sometimes the report reflected some recent event rather than the entire period." The bias from the most recent performance was corroborated by comments from the Council Of Defense And Space Industry Associations. "CPARS assessments often contain outdated

information or are focused on the "issue du-jour" at the expense of underlying trend or longer term performance assessments. (2013, pg. 5)" Another informant implied that the government's CPARs are inflated, stating: "several primes have their own rating system for suppliers. My perception is that they tend to give lower scores."

All informants mentioned inconsistency and subjectivity in the CPAR evaluations. One informant commented: "Inconsistency given by the human judgment factor. There is too much subjectivity." "It makes reports unreliable." Another informant stated that the CPAR was not fair due to too much subjectivity, and reported experiencing inconsistent definitions of the ratings across CPARs. Another informant commented: "In our experience [the agency] does not follow and blatantly violates published guidelines for filling these out and uses subjectivity to cover up for its own mistakes during project execution." One informant commented how the customer can be internally inconsistent: "When a quarterly review with the customer comes back four quarters in a row with an exceptional /very good write up its very hard to accept a satisfactory at the end of the performance period." Another informant commented: "With the CPARS, I know there's supposed to be some guidelines on, you know, what's acceptable or marginal, or whatever the guidelines are for performance, but it seems to be not really strictly enforced as far as like there's more motion on the ratings of the CPARS with not a lot of justification." This comment suggests that inconsistency is attributed to the individual evaluator conducting the evaluation: "We've had some difficult people leave, so we had some new people come in and things got a lot better and then we've had it go the other way." One informant stated: "There are some shortfalls in it, it is definitely subjective. In other words, most especially between different customers - meaning that we have contracts with various contracting agencies across the government. Some are a lot harder raters than others." Another informant commented: "Now above that when you are talking about satisfactory to very good and exceptional performance, much more subjective, much less of a leg to stand on." He went on to say:

> "there is a big difference between the way different commands evaluate or score CPARS. I have got one that consistently if you are basically doing the job it is satisfactory, you are green. They might give a few very goods. It really takes something strong to get a very good. I have never seen an exceptional from them. Other ones very goods are the norm and the exceptional are not uncommon and very goods you have to be basically doing some things wrong—again it is not to a marginal category, but that is where it gets very subjective. If you miss a [data deliverable] or two, well what does that mean? Does that mean that you are yellow because you missed some? Well it shouldn't because how

> important were those and what was the circumstances and such? Then that is just depending on who is making the judgment how [can] we score."

Another informant corroborated the problem.

> "I see the difference are in the relative ranking, in other words, a program that has all greens and three purples can be just as good as a program that has exceptionals with a couple of very goods, but they are judged from different contracting offices and one puts real high standards on what scores—what it takes to be considered above satisfactory and the other kind of lets loose as far as how people judge things."

Timeliness of CPAR reporting was identified as a concern: One informant commented, "We have had reports over a year late." Another reported: "the contractual period of performance ended 31 December. I just received the CPAR from the CPAR system 11 April. The previous period of performance machinations lasted until June. 6 months after the end of the period."

Timeliness of performance feedback in general also seems to be a culprit. When asked whether the government uses CPARs to actively manage contractor performance, one informant commented: "Don't think they do." A strong consensus was that CPARs are not used to manage contractor performance on an ongoing basis. One informant commented: "It is more—it is easier to commonly hear them, like "Oh, CPARS again." Again, you know, just with appraisals, "Oh, appraisal time again." Yeah, they don't look forward to it and so I think for them it is not a priority, it is not a means of measuring." Another informant stated: "There is not a whole of other performance feedbacks throughout the year." One informant stated: "can't very well do a better job if you aren't informed until the end of the year." Another informant alluded to the importance of time stating: "We have been more proactive with our performance evaluation discussions with the client through the performance itself, so we don't get blind [sided] like the one bad experience we had." One informant stated: "when we had more award fee contracts, we got more, you know, verbal feedback or specific feedback. Once we went to incentive fee contracts, we don't get that. I would rather, you know, more times through the years, someone at the [customer] gets some comments as to contractor performance rather than waiting until the CPARS." One informant commented on the value of more frequent feedback intervals: "I have got one contract for sure that there is a monthly scorecard. It is the—the government gives us a monthly look at

how we are doing and it basically very much you can tie it right to the CPAR. There is no fuzz on how things are going throughout the year. So that is great."

Poorly defined requirements and differences in expectations were identified as weaknesses that affect the accuracy of SPEs. One informant stated: "Absolutely. It is difficult to meet an "unknown" requirement." Another informant mentioned a lack of agreement (i.e., not in the contract) on performance levels, stating: "And if you say you want a [inaudible] cap level of a certain point, yeah, we can make that, but you have to agree that that's what our threshold is going to be." "Sometimes it is [in the contract], but they've gotten rid of our incentive fee, so when we don't have an incentive fee threshold and you don't write it into a contract or a performance work statement, we're kind of shooting in the dark." "Just tell us what it takes—what you expect. If we want to get an exceptional, what do you expect us to provide to you? What performance?" This statement underscores frustration with the ambiguity in what performance level is required to attain above-satisfactory ratings. Some informants mentioned buyer-side expectations that were not captured in the contract: "they're trying to hold us accountable for things that aren't even in the contract." "What they wanted and what they bought were two different things. And so I'm getting dinged on things that again I shouldn't be dinged on. It was not in my contract to provide that level of service for talent." One informant raised the issue of using the wrong contract type, how that affected differences of expectations of performance, and ultimately affected the resultant CPAR.

> "a customer needed a replacement part and it was obsolete. We needed to redesign another replacement and the—from the customer standpoint it seemed pretty simple. "Hey, this part is no longer being made, we can no longer get it. [Contractor], we need to you create a new one. Can you get it for us?" It is not that simple. In other words, trying to resurrect what went into this design of some piece that is 20 or more years old and work with the subcontractor who is maybe turned over the ownership of the company a couple of times and such, that there is no doubt that it is [the contractor's] need to fix it, but just a recognition on both sides about too simplistic on what the fix would be, that we realized that it is a much harder job. Now that—yes, [the contractor] is on the hook to do it, but it had to do from both sides they just didn't realize what the real requirements were up front or what was all going to be needed to be done. It goes to the idea of not accepting fixed price developmental contracts. Basically trying to hold our feet to a fixed price contract on the developmental side when that is definitely the wrong contract vehicle to do it under."

Evaluator turnover also surfaced as a contributing factor. When asked whether turnover affects past performance evaluations, one informant commented: "Very much so. New players may not be aware of particular circumstances leading to performance issues. Or like in one of our well established programs, the PCO changed after 10 years. And since the program was always on time and within budget our CPARs were rated mostly excellent (4's and 5's) but the new PCO said she never gave better than 3. So to an independent reviewer our performance looked as if it went down when in fact it was the same." Another informant commented: "on my current contract, I'm on the third PCO since I took the job over about a year ago."

One informant commented, "There are a few cases where it just seemed like somebody had an agenda. Those never go over well." Another informant corroborated the existence of an agenda stating: "[The evaluator] distorted the evaluation to suit his/her own agenda." These testimonies suggests that CPARs are sometimes used to attain some purpose other than to provide the supplier honest, accurate performance feedback. A third informant commented: "The CPAR process from the Contractor side can be brutal. The Government personnel can abuse the system and mete out punishment with little to no recourse. The Contractor always appears to be in the wrong as they are replying to accusations."

One informant identified inadequate training as a culprit. "See major weaknesses in the training government employees receive on criteria to use in filling these out. This has cost us tens of millions of dollars in business on future bids."

One informant aluded to a lack of due diligence on behalf of the customer, stating: "Lack of knowledge on the Customer side. I have been evaluated for performance outside the period of performance. Additionally, I have been evaluated by Government employees who do not research their comments before entering them into the CPAR system. The Government has continued to ignore contract language to down grade our performance rating." Another informant commented:

> "It depends on the COR. Some are very thorough and some are not and especially the one that wasn't was the one I had the issue with and then went back for verification. "I am not understanding," or, "you're interpreting it one way and they are interpreting it a different way." So, yes, it differs. Then you have some you can tell—again just like performance reviews, you can tell when the COR actually took his or her time to do it and then you can tell—I have had ones where you can tell that they literally took five minutes to do it and [inaudible] last minute."

Some informants identified a lack of justification for the assigned ratings. "The justification for ratings received have been weak to minimal on the Government's part."

One informant identified the function of the evaluator as a factor. "the closer you are to the product, you know, the less disagreement we have." The informant was aluding to more disagreement with contracting personnel than with program managers and engineers with whom they interact more often.

The following exchange serves as quite a vivid example of the failure of communication as a contributing factor to inaccurate SPEs.

> "I've had three different task orders that have been awarded to us under this IDIQ and two of them because, you know, money can be [inaudible] in these interagencies that can use big large idea vehicle. For the [X] contract, two of my [inaudible] do come from [X], but one actually comes from [Y] and so it's interesting seeing how the different agencies handle it even under the one umbrella permit. So I would say the [X] guys are very much more work with the contractor in the sense of at the end of the day there's no surprises in my CPAR. You know, at the point I address it, we've had open communications along the way, there's a lot of back and forth in them considering my input, and that's worked well. On the other side, on the [Y] side, so as I mentioned, it could be in their policy, it could be their environment or training, but I was honestly pretty surprised by the CPAR that I got and I could literally go in and justify it all the way to the issue that I [inaudible] down in the evaluation, but it wasn't until the end, you know. I sent a CPAR, then they put it in, and I have some, you know, so long to respond, and then I made my response. The COR didn't agree, so it went up to a higher up in the [Y] and then the actual higher up person in [Y] went back and actually cited some certain things that were out of the contractual plans. One was we got dinged on key or COR personnel with no key or COR personnel requirement in my contract, so I had something to say about that. I'm not sure why I'm being rated on this when there are no COR or key personnel requirement in the contract or in its deliverable. Obviously, I decided it was in my favor to remove that. It was almost afraid to talk to me. You know, it was like they were afraid of the contract or afraid of what they can or cannot say, so it was very different from my other experiences. Now, [that was] like in our base period the

first year and now this time it's funny when he delivered one, he's like, "Yeah, we've got this CPAR to do again." He's like, "I've been told I can actually get your input."

This supplier also identified a lack of training of assessing officials: "I think in the CPAR process the greatest weakness would probably be training for the COR or COTR."

One informant reported engaging and appealing to a higher-level authority in the buying agency to rectify an inaccurate report.

One informant mentioned that past government errors or failures that were erroneously attributed to the contractor – discovered after the close of the CPARs – never get reflected in the CPAR retroactively. The report, as it was, remains inaccurate in the system.

**P24: Suppliers' reputation preservation efforts will be positively related to SPE rating inflation (i.e., negatively related to accuracy).**

**P25: Inconsistency of rating definitions will be negatively associated with SPE accuracy.**

**P26: Subjectivity of SPEs will be negatively associated with SPE accuracy.**

**P27: Feedback frequency moderates the relationship between inconsistency and SPE accuracy such that more frequent performance feedback decreases the magnitude of the negative relationship between inconsistency and SPE accuracy.**

**P28: Feedback frequency moderates the relationship between subjectivity and SPE accuracy such that more frequent performance feedback decreases the magnitude of the negative relationship between subjectivity and SPE accuracy.**

**P29: Performance feedback frequency will be positively related to SPE accuracy.**

**P30: The sufficiency of the requirement definition will be positively related to SPE accuracy.**

**P31: The greater the assessor turnover, the lower the SPE accuracy.**

**P32: There will be a positive relationship between the assessor's level of effort and SPE accuracy.**

**P33: Assessors serving in functional capacities with less knowledge of the technical aspects of supplier performance will generate less accurate SPEs.**

**P34: Assessor training on SPEs will be positively related to SPE accuracy.**

**P35: Assessor level of effort will be positively related to SPE specificity.**

Together, the aforementioned propositions are displayed in Figure 5.

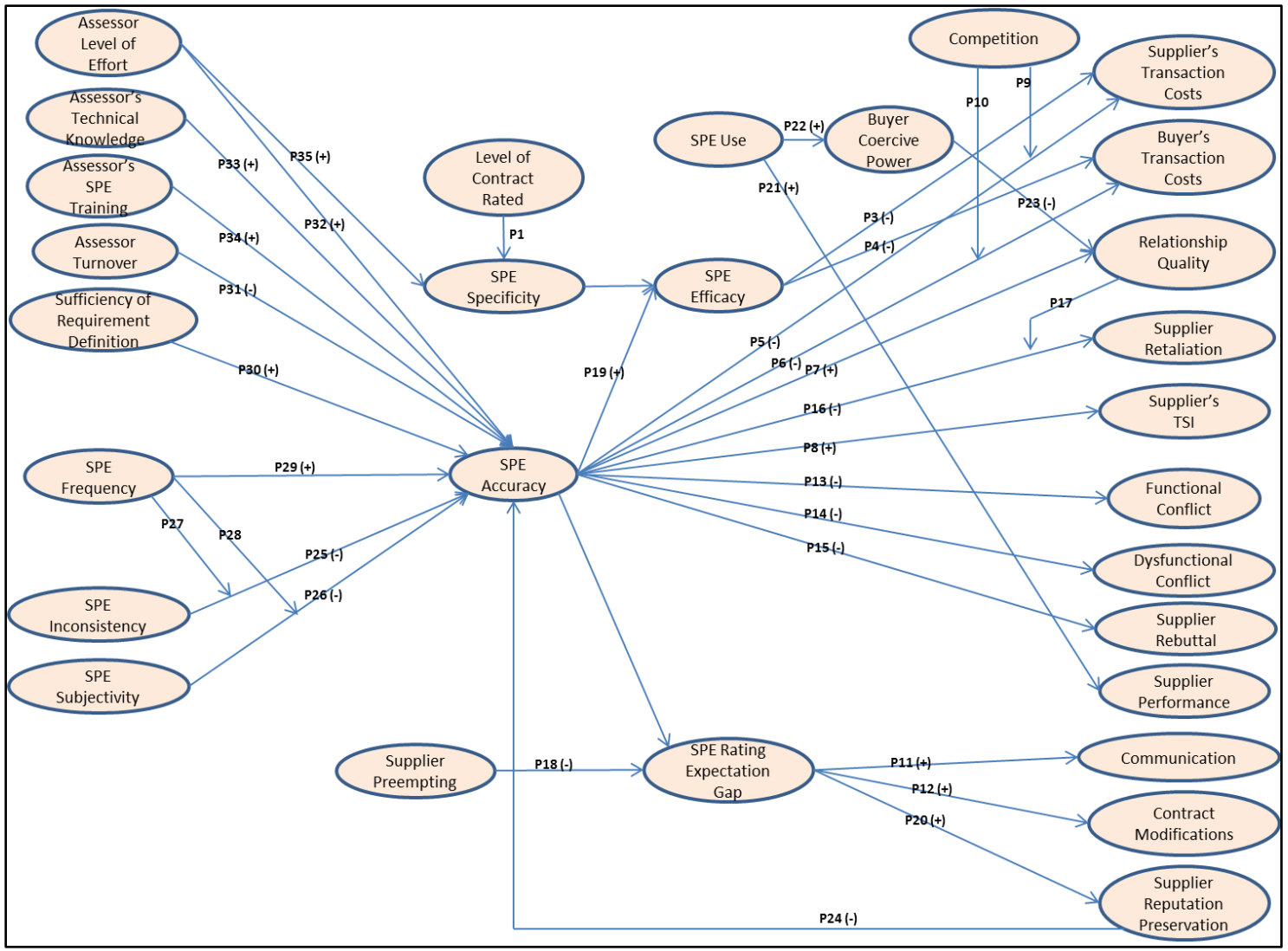


Figure 5. Consequences of SPE Efficacy (Supplier

### 6. How many man-hours do suppliers invest in responding to SPEs?

Across the eight respondents, reported man-hours varied widely. Some informants indicated that the time depended on whether their firm agreed with the CPAR evaluation or whether the evaluation was positive or negative. If positive, man-hours were minimal (one estimate provided was two man-hours). If, however, there was disagreement (or negative), man-hours consumed to respond to the government buyer ranged from 15-800 (average 202, median 80). One informant attributed the extra effort to "all of the management involved." Another informant commented: "And as you look at over like the 20 IPTs I go to, it takes 10 or 15 man hours and it starts to add up. It's not efficient for them or us." "A lot of these are billable hours." "I think that it does cost each one too much money for something that should not be this hard to do." "I think that because it has become so huge, they'd say, 'Yeah, we're going to just go ahead and bill because it takes us so much effort to do this that we might not have otherwise." "It's not effective for the person who has to go out and fly the aircraft or the person who has to maintain the aircraft. It does them no good. We spend dollars on unimportant administrative paperwork."

### 7. What communication tactics do suppliers use to manage the SPE process?

Tactics used by suppliers to manage the SPE process varied from no attempt to manage the process to very deliberate efforts. One informant reported providing quarterly inputs of performance to the government customer coupled by quarterly reviews by the customer. This contractor also provides the customer a draft CPAR report prior to the end of the period of performance. "We decided that [draft reports] gave better feedback and that if there was an issue, we could adjust it during the year, which has been suggested by our company that that would be a good way to go it and it would stem off issues long before the end of the year." The contractor cautioned, however, that the draft report has been met with some resistance by the customer, who perceived that the contractor was trying to tell the customer how to do their evaluation. Another contractor also preempts the official CPAR with a draft self-assessment provided to the government six months in advance of the due date for the official CPAR. This contractor commented: "It gives us a much better leg to stand on and a much better negotiating position to utilize the self-assessment." The self-assessment approach seemed to pay off. "I really think since we started the self-assessment process that the number of CPARS that we need to rebut is much smaller. I think maybe it used to be 10% and now it is down to 5%." Another informant tries to preempt the CPAR informally, for example at the end of a meeting such as a critical design review.

The informant using the self-assessments also mentioned that it is helpful to, on a large program with many layers of management, ensure that the self-assessment is coordinated with all of the customer's functional performance assessors at the lowest level. "So I mean the lesson learned that way, we need to get those self-assessments down to the focal level."

One contractor commented: "There is not a whole of other [than CPAR] performance feedbacks throughout the year." One informant reported providing "just constant, constant communication…keeping [the customer] aware," after being surprised by a less-than-favorable CPAR. A common thread in responses pertained to increased communication and more frequent communication.

Another informant mentioned using self-praise to call the buyer's attention to things done well: "You know, giving them updates of even when I think something is trivial [inaudible], you know, the kudos, you know, patting yourself on the backside."

Interestingly, when asked whether the CPAR process and/or rating definitions were discussed at the post-award orientation, one informant commented: "No, never." No other informant mentioned CPARs as part of the post-award orientation agenda.

One informant mentioned increasing documentation: "We will make sure that is well documented and [inaudible] that I don't get dinged on that on in the CPAR."

# Discussion

Substantial transaction costs are dedicated to avoid adverse selection - the risk of selecting an incapable supplier that otherwise misrepresents itself as capable. Following contract formation, more transaction costs are incurred to monitor supplier performance to thwart supplier opportunism ex post. The effectiveness of a mechanism to monitor and record supplier performance information, a supplier performance evaluation, was the topic of this study.

There are many concerns that the SPEs/ratings are not properly, timely, or accurately completed. Unreliable or inaccurate past performance assessments can harm suppliers' reputations and can bias source selections, resulting in adverse selection. If past performance information is not reliable, and if 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 would be wasted. Anecdotal evidence suggests that buying organizations often do not use past performance information as a meaningful discriminator between proposals.

The purpose of the research, therefore, was to explore the antecedents to and consequences of the efficacy of SPEs. The intent was to diagnose alleged weaknesses and to explore potential improvements. The following research questions were addressed:

1. What factors decrease the efficacy of SPEs?
2. How do suppliers react to inaccurate SPEs?
3. Do SPEs, in general, motivate suppliers to increase performance?
4. How does the accuracy of SPEs affect relationship quality?
5. Why are SPEs often inaccurate?
6. How many man-hours do suppliers invest in responding to SPEs?
7. What communication tactics do suppliers use to manage the SPE process?
8. To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy?

This research combined quantitative and qualitative methodologies to examine these research questions. From a literature review, a conceptual model of 27 hypotheses was developed to explore antecedents to SPE efficacy. A survey was deployed, and data was collected from 131 respondents. The model was

analyzed using PLS SEM to estimate relationships. To explore the consequences of SPE efficacy on suppliers, a qualitative approach was employed. Eight subject matter experts representing suppliers were interviewed to explore how well SPEs achieve the intended goals of: (1) motivating supplier performance and (2) decreasing the risk of adverse selection in the future. The data analysis resulted in the development of 35 testable propositions that should yield insight into the phenomenon from the supplier side of the dyad.

1. What factors decrease the efficacy of SPEs?

While most respondents agreed that SPEs are effective, an appreciable number - nearly a third - challenged the effectiveness of SPEs. Figure 6 depicts the series of relationships that explain SPE efficacy. From the survey of buyers, SPE efficacy was deteriorated directly by rating dissonance and by poor rating justifications. Additionally, some evidence suggests that low SPE efficacy is associated with inflated ratings. The effect size of rating justification on SPE efficacy is more than twice that of rating dissonance. Therefore, buying organizations needing to improve SPE efficacy should first seek means to improve rating justifications. Insights as to how to do so follow.

Perceived usefulness of the SPE process and supporting information technology tools and the perceived accuracy of SPEs affect how sufficiently SPE ratings are justified. Thus, performance evaluators with little faith in the SPE process(es) and tool(s) may not invest the requisite time and effort to justify their ratings. Hence, it appears that in a manual SPE schemes that depend largely on human effort and discretion, evaluators must believe in the SPE scheme's efficacy; otherwise, rating justifications will suffer. Perceived accuracy has a far greater effect on rating justification than does perceived usefulness. Role overload also decreases rating justifications. Further, perceived accuracy also affects perceived usefulness. Examining the effects on perceived accuracy, the sufficiency of the requirement definition has the greatest impact, followed closely by buyer fairness. Leverage attitude slightly decreases perceived accuracy, while communication bi-directionality and communication formality increase accuracy. Contrary to conventional wisdom, the perceived accuracy of SPEs does not improve with increased surveillance of the supplier. Some evidence also suggests that accuracy is affected by inflated ratings.

Rating dissonance is increased by fear of supplier dispute and by evaluator role overload. Fear of supplier dispute increases rating inflation, and is decreased by relationship quality. Relationship quality is degraded by evaluator turnover and enhanced by communication bi-directionality and communication formality.

Rating dissonance was not affected directly by perceived accuracy. However, an indirect effect of perceived accuracy through fear of supplier dispute exists. Thus, lower accuracy increases fear of supplier dispute, which, in turn, increases rating dissonance. Some evidence suggests an association between rating dissonance and rating inflation. Rating dissonance is also increased by excessive amounts of workload (i.e., role overload). Evidence suggests that the SPEs examined in the sample were manual processes of information collecting, synthesizing, deliberating, and reporting supplier performance. Greater perceived accuracy in SPEs resulted in more time to complete the SPEs. Since the SPEs examined in the sample were largely manual processes, over-work situations can result in evaluators sub-optimizing tasks that are perceived as less critical or tasks in which a minimalist approach can go undetected (perhaps tasks such as SPEs).

SPE efficacy was most affected by the extent to which ratings were justified, documented, explained, and understandable. While SPE efficacy was not affected directly by perceived accuracy, there was an indirect effect through the effect of perceived accuracy on rating justification. Thus, SPEs that are less accurate do not sufficiently justify the ratings. Additionally, post hoc tests unveiled another potential culprit to poor justifications – poor rating definitions. In turn, both stakeholders of SPEs are affected. Future buyers are affected during source selection by struggling to understand the ratings and how to evaluate them. Suppliers are also affected. It may be that in order for suppliers to allocate more resources and effort into attaining higher levels of performance, they have to believe the ratings. Thorough rating justifications may help them to buy in, while poor justifications may spawn no action.

Lower SPE efficacy also appears to be associated with inflated ratings. Suppliers suggested that the level at which SPEs are reported affects the efficacy. SPEs completed at a parent contract level (e.g., IDIQ) versus a task-order level are not effective; they are not specific enough for future source selection teams to discern relevancy or to glean insights on performance.

Some buyer-side hypothesis testing conflicted with supplier-side propositions. For example, suppliers' experiences suggested the existence of relationships between assessor turnover and SPE accuracy. However, the data collected from assessors did not support this hypothesis. These results could be attributed to the source of the data; thus, more research is needed that combines data from the buyer and supplier of the same transaction. The results also suggest that there may be gaps in perceptions of accuracy between buyers and suppliers. Suppliers contested SPEs 25 times challenging 45 ratings.

It appears that suppliers commit a significantly greater amount of time reacting to SPEs (mean 202 hours) than buyers do preparing them (mean 18.1

hours). However, additional research based on a larger sample is needed to confirm the suppliers' estimates.

Some evaluators adopt an attitude that using SPEs as leverage is legitimate practice. This leverage attitude is associated with lower SPE accuracy. Attitudes toward leverage were higher than actual leverage employed. Testimony from suppliers confirmed that evaluators sometimes use SPEs to exercise coercive power. This research also unveiled a distinction between proactive leverage (getting a supplier to do something) and quid pro quo leverage (receiving payback for an inflated rating), with the former being more common.

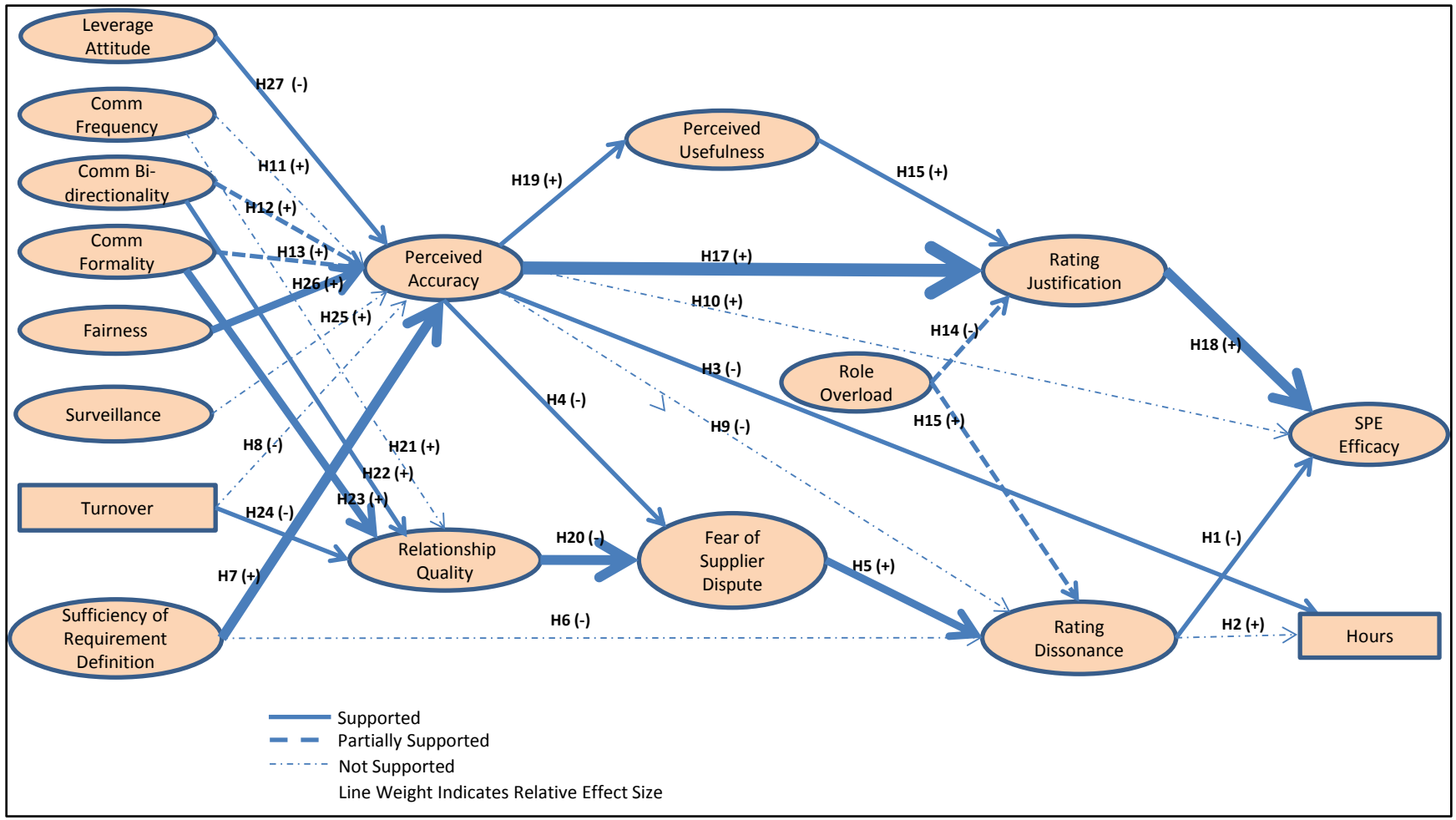


Figure 6. Antecedents of SPE Efficacy

2. How do suppliers react to inaccurate SPEs?

When receiving a SPE thought to be inaccurate, suppliers work to correct the record. They incur significant transaction costs in doing so. SPEs thought to be inaccurate lead to functional and dysfunctional conflict. Functional conflict encompasses the parties working together to understand the ratings, the rationales, and corrective actions (perhaps on both sides of the dyad). Several suppliers reported making a concerted effort to understand the customer's positions, then taking added measures to better define the customer's performance expectations. Then, suppliers reported taking measures to ensure those newly refined expectations will be met in the future. Dysfunction conflict manifests in destructive actions such as blame, dispute, or retaliation. The buyer-supplier relationship also appears to suffer with inaccurate SPEs. Most suppliers reacted by rebutting the assessment in cases where they believed their rating and/or justification was not accurate. In doing so, several suppliers reported involving senior management in the reaction process. Some suppliers also reported that their company has rebutted CPARs whose ratings were negative even when they believed the CPARs ratings were warranted. Thus, regardless of the truth, some suppliers will take measures to preserve their reputation. In addition to accuracy, suppliers suggested that the gap between the expected and actual rating is key. Hence, even positive ratings (e.g., "acceptable" or "good") may be unwelcome and, thus, disputed. One supplier reported abandoning a customer permanently. Suppliers also mentioned an unwillingness to go "above and beyond" for a customer following an inaccurate SPE; they will revert to providing the bare minimum stated in the four corners of the contract. At face value, this might be dismissed by some buyers; however, considering that many requirements are ill-defined (Hawkins et al., 2015) and that complex contracts are always incomplete (Williamson, 2005), depending solely on the contract may be detrimental to the buyer. Another informant reported an increase in a willingness to bill for administrative hours spent correcting the CPAR. Thus, some evidence suggests that suppliers retaliate or attempt to get even in some way. Some suppliers also reacted by preempting later CPARs processes by providing performance rating input to the assessing official prior to the CPAR due date. One informant mentioned increasing documentation will increase transaction costs.

Suppliers raised two particular consequences of SPE efficacy – buyer and seller transaction costs. Rather than focus on SPE efficacy – a buyer-centric concern, suppliers discussed SPE accuracy. They described several deleterious effects on the supplier of low accuracy, and discussed sources of inaccuracy. Their testimonies are largely consistent with the findings from the buyer-side survey data

that rating dissonance likely traces back to inconsistent interpretations of ratings and performance criteria coupled with subjectivity.

3. Do SPEs, in general, motivate suppliers to increase performance?

The responses to this question were mixed; however, evidence suggests that SPEs can be an effective motivator. One informant qualified the effect contingent on the buyer's proper use of the evaluation processes. One informant clearly backed the relationship between SPEs and an attempt at attaining higher performance. However, three informants reported no effect on their level of effort; they strive for excellence regardless. One informant attempted to increase performance to the extent that they could avoid receiving marginal or unacceptable ratings. Thus, further research is needed to understand the conditions in which SPEs do not motivate higher performance.

4. How does the accuracy of SPEs affect relationship quality?

One supplier informant reported that his company will no longer do business with a particular government organization due to misplaced blame on the contractor for repair delays; thus, in this case, the relationship was ruined. Another informant identified a strained relationship with deteriorated trust. While the supplier side of the dyad is the most appropriate perspective to take in determining the effect of SPE accuracy on relationship quality, buyer-side data corroborates a positive direct effect. Relationship quality in this data omitted commitment; it encompassed only trust and satisfaction.

5. Why are SPEs often inaccurate?

Accuracy of SPEs is critical because it decreases supplier rebuttals and, indirectly, increases SPE efficacy. Several factors affect perceived accuracy. From the buyer survey data, an insufficiently defined requirement has the greatest impact on accuracy. Suppliers corroborated that poorly defined requirements and differences in expectations were identified as weaknesses that affect the accuracy of SPEs. Therefore, buyers and suppliers should work to better define requirements and how fulfillment of those requirements will be verified. These details should be sufficiently explicated in the solicitation and contract documents. Unfairness by the buyer also decreases accuracy. Hence, the closer the SPE resembles what the evaluator discerns the supplier deserved (i.e., earned), the greater the perceived accuracy. If the SPE deviates from that deserved – positively or negatively, accuracy decreases. Thus, evaluators should record SPEs that reflect only what the supplier earned. Hence, sources of distortion (see Figure 2) should be purged. Explicit policy and process automation could be helpful in this

regard. Leverage attitude slightly decreases perceived accuracy. Suppliers corroborated the effect of leverage, reporting a hidden agenda in some SPEs. Some evidence suggests that a lack of communication formality and communication bi-directionality degrade accuracy. Thus, process and policy changes and any augmenting information technology should facilitate formal and two-way communication. Evidence from suppliers (and qualitatively from some buyers) suggests that infrequent performance evaluations decrease accuracy. However, general communication frequency did not affect accuracy in the sample. An automated supplier management information technology system could facilitate more frequent SPEs. Both the quantitative evidence from buyers and qualitative testimony from suppliers suggest that inflated ratings are associated with lower perceived accuracy. Thus, mechanisms to thwart rating inflation should be explored, such as instilling means to reduce subjectivity in evaluations. Suppliers identified biases commonly mentioned in employee performance appraisal literature – recency and an emphasis on the negative. Suppliers also mentioned inconsistency and subjectivity in the CPAR evaluations as the greatest contributors to inaccuracy. Although not supported by the buyer data, suppliers identified evaluator turnover and technical knowledge as factors that decrease accuracy. Suppliers also mentioned that a lack of training of evaluators on how to conduct a SPE was a factor.

6. How many man-hours do suppliers invest in responding to SPEs?

Across the eight respondents, man-hours varied widely. Some informants indicated that the time depended on whether their firm agreed with the CPAR evaluation or whether the evaluation was positive or negative. If positive, man-hours were minimal. If, however, there was disagreement (or negative), man-hours consumed to respond to the government buyer ranged from 15-800 (average 202, median 80). These hours constitute significant transaction costs. Nonetheless, the magnitude across all rated contracts that are contested is unknown. Future research could seek to quantify the transaction costs from a representative sample. Given the suppliers' testimonies, greater accuracy of SPEs should result in fewer disagreements and rebuttals, which should, in turn, reduce transaction costs.

7. What communication tactics do suppliers use to manage the SPE process?

Tactics used by suppliers to manage the SPE process varied from no attempt to manage the process to very deliberate efforts. One informant reported providing quarterly inputs of performance to the government customer coupled by quarterly reviews by the customer. This contractor also provides the customer a draft SPE report prior to the end of the period of performance. One informant

mentioned that it is helpful to, on a large program with many layers of management, ensure that the self-assessment is coordinated with all of the customer's functional performance assessors at the lowest level. A common thread in responses pertained to increased communication and more frequent communication. No informants mentioned SPEs as part of the post-award orientation agenda. One informant mentioned increasing documentation.

8. To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy?

Rating dissonance decreases SPE efficacy, but not very strongly. Dissonance is linked to evaluator role overload and is strongly affected by fear of supplier dispute, which is increased by lower SPE accuracy. Some evidence suggests that dissonance is increased by inflated ratings.

## Managerial Implications

Buy-side evaluators overall reported that SPEs are somewhat effective in mitigating the risk of future adverse selection, and they were rather ambivalent as to whether SPEs motivate suppliers to perform. However, some suppliers interviewed reported that SPEs indeed motivate them to perform well. In addition to asking respondents to rate the efficacy of their most-recently completed SPE, respondents were asked to rate the extent to which he or she relied on past performance in making an award decision in his or her most recent source selection. The average midland response (4.34 on a 7-point scale) raises questions as to the extent that SPEs reduce the risk of future adverse selection.

In examining SPE efficacy, several novel factors emerged. For example, some main findings centered around the dissonance among multiple performance evaluators on a single contract. Another major finding entailed the importance of a justified and accurate rating. These constructs significantly affect SPE efficacy either directly or indirectly. The findings herein introduce a plethora of implications for supplier management, discussion of which follows.

Rating dissonance reflected the extent to which raters struggled to reach consensus on the supplier's rating(s) and/or narrative justification. Overall, the magnitude of rating dissonance was not high. Nonetheless, rating dissonance decreases SPE efficacy. Rating dissonance did not relate directly to a sufficient definition of the work to be performed by the contractor. Neither was rating dissonance related directly to the accuracy of the SPE. Rating dissonance may be attributed to a lack of a common meaning of performance criteria and of rating definitions. Looking at the post hoc tests and at Appendices D (evaluators'

recommended policy changes) and E (reasons for dissonance), it is apparent that criteria and rating meanings are not well defined by the buying team ex ante.

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 (staffing) 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 SPEs. This research revealed that, on average, SPEs consume nearly two days of effort by buyers to conduct the SPE and two weeks by suppliers to react to them. These transaction costs should be extrapolated over all contracts requiring a SPE to determine total transaction costs of using a manual SPE policy and reporting system. Next, alternatives could be explored to determine whether commercially available supplier management systems could reduce total costs and improve accuracy and SPE efficacy simultaneously. Hence, can commercially available systems automate any of the performance measurement and SPE reporting processes such that transaction costs could be decreased? And, can a commercially-available system improve the variables that affect SPE efficacy such as accuracy, rating inflation, communication, workload, fairness, sufficiency of the requirement definition, perceived usefulness, rating justification, sufficient rating definitions, sufficient performance criteria definitions, links between performance levels and assigned SPE ratings, relationship quality, fear of a supplier dispute, and rating dissonance?

The greatest factor determining SPE efficacy was rating justification. Thus, for those seeking to increase SPE efficacy, efforts should be made to more thoroughly justify ratings. This research offers insights as to how to improve rating justifications. First, buying organizations can implement information technology systems that are useful. Additionally, buying organizations can address the effort required to justify a SPE. This can be done by making more time available to evaluators to conduct SPEs by hiring more evaluators, by dedicating evaluators to the task of supplier performance evaluation, or by reducing evaluators' non-SPE duties. Rather than addressing workload capacity, organizations can seek to reduce the amount of effort required to produce a sufficiently justified rating. The most logical means would be via process automation (e.g., a supplier management information technology system). Some evidence suggests that buying organizations can improve rating justifications by sufficiently defining rating definitions. To do so, it may be necessary to tailor ratings and their definitions to the particular goods or services being procured. Sourcing teams should further define performance criteria, how each will be measured, and develop thresholds for each that unambiguously lead to the specific performance ratings. These

performance criteria and rating definitions should be defined in the request for proposals and requirements documents, and then set in the resulting contract. Consequently, the supplier would know precisely how its levels of actual performance will translate into performance ratings, and it will more likely believe and accept the ratings as legitimate. In turn, this research suggests that SPEs will more effectively motive the supplier to increase performance and will better inform future source selections (i.e., reduce the risk of adverse selection). A commercially available supplier management system may prove useful as it would instill structure into the measurement and calculation of performance levels and help automatically translate performance into SPE ratings. Therefore, those organizations that do not use structured SPM applications should consider doing so. Buying organizations could also develop a SPE quality index as a means to periodically audit SPE quality.

A central construct affecting SPE efficacy appears to be the accuracy of the evaluations. In fact, lower accuracy is directly related to supplier rebuttals. This is consistent with the literature. "Strategic sourcing is not possible without tools for measurement, analysis and follow-up of the category. Without data, the work changes from being problem-solving based on facts to being a debate about opinions" (Carlsson, 2015, p. 126). While the overall sentiment in the sample was that SPEs are somewhat accurate, of all constructs measured, accuracy varied the most. Accuracy was found to be affected by communication bi-directionality, communication formality, buyer fairness, leverage attitude, and sufficiency of the requirement definition. Accuracy of SPEs 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, buying teams should not move forward in contracting with ill-defined requirements. Additionally, prospective suppliers should strive to ensure that the buyer thoroughly defines its requirements. For services, an independent requirements ombudsman could help in this regard. Some evidence suggests that accuracy is degraded by rating inflation.

SPE efficacy is not the only consequence at stake. For-profit buying organizations often use SPEs to rank suppliers then reward or penalize them according to the ranking. The DoD recently implemented such a program dubbed the superior supplier incentive program (USD AT&L, 2015). If any of the underlying SPE and supporting performance data is inaccurate, the validity of rankings may be suspect. Ultimately, an undeserving supplier receiving benefits and recognition may displace a deserving supplier.

This research highlights the limitation of relying on largely manual SPE policies, procedures, and supporting information technology (e.g., CPARS). In the federal government, there is no single structured information technology system

and process to systematically collect, store, and synthesize supplier performance information. Yet supplier performance management systems are common in the for-profit sector. Examples include Iasta'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, rating determination, and reporting. They also offer dashboard-like scorecards to assess individual suppliers and common groups of suppliers (e.g., by commodity family or by industry).

Such a structured tool could alleviate many of the weaknesses that deteriorate SPE accuracy, facilitate inadequate rating justifications, and accommodate rater dissonance, while bolstering the ability to manage suppliers' performance levels. Such a tool could reduce or eliminate the many instances of incomplete ratings (e.g., rating categories not rated) by, as shown in this research, improving perceived accuracy, perceived usefulness, and communication bi-directionality. Automated, more fact-based data collection and use could also reduce the instances of changes to SPEs by higher authorities (e.g., "reviewing officials"), either due to need for correction or due to human manipulation to attain alternative goals. Automation would also enable more frequent SPE feedback to the supplier and, thus, reduce recency bias that can plague long-interval evaluations. More fact-based evaluations should also alleviate the need for evaluators to solicit suppliers to write their own evaluations – as was common in the sample. The value of non-independently-derived performance information is suspect. It may result only in the supplier's opportunity to provide marketing material, increases the supplier's transaction costs, and could be billable to the customer under a non-fixed-price-type contract. Policy could also address whether the practice of suppliers writing their own performance evaluations is allowable.

Automation could also reduce transaction costs required to conduct SPEs. In fiscal year 2015 alone, 2,228,275 SPEs were either conducted, in-process, or required to be conducted (i.e., overdue) (Naval Sea Logistics Center Portsmouth, 2015). Assuming: (1) a consistent number of SPEs annually, (2) a rate of pay of government evaluators and contractor employees equivalent to a GS-13, step 5, (3) a fringe benefit rate of 36.25% (OMB, 2008), (4) that 19% of SPEs will be rebutted, (4) that contractors spend 2 hours on non-rebutted SPEs and 80 hours on rebutted SPE, and (5) that buyers spend 8 hours on each SPE – each as found in this research (medians), SPEs will require the full effort of 26,512 full-time equivalents and cost $2.99 billion annually.

This research confirms a halo effect (i.e., rating inflation) attributed to a fear of supplier dispute. This research corroborates previous anecdotal reports that

evaluators and reviewing officials change (i.e., increase) ratings in order to: (1) avoid conflict, (2) protect a program, (3) preserve the supplier relationship, (4) gain leverage over the supplier, and (4) avoid harming a supplier's future business opportunities. More objective performance measures explicitly linked to precisely defined ratings could be used to increase supplier buy-in to ratings, thereby reducing fear of supplier dispute and rating inflation.

Some performance evaluators believe it acceptable to use the SPE rating as leverage—either (or both) as a threat to a supplier during performance and prior to a SPE or as a means to extract concessions post hoc from a supplier in exchange for a more favorable SPE rating. However, few respondents acted on those attitudes. The survey data from performance evaluators was corroborated by testimonies from suppliers. Such an attitude toward leverage was marginally related to lower SPE accuracy and had a small effect. Nonetheless, since accuracy was found to be a central construct leading to SPE efficacy, unintended uses of SPEs as leverage should be explicitly addressed in training and policy.

Suppliers questioned the utility and accuracy of SPEs that are conducted at a parent-contract level (e.g., IDIQ) versus a task-order level. This practice should be explored further to determine the extent to which higher-level reporting hinders SPE efficacy (i.e., motivating supplier performance and mitigating the risk of future adverse selection by informing future buyers).

This research highlights the benefits of building and maintaining quality relationships – those characterized by high trust and buyer satisfaction. In the SPE process, relationship quality decreases supplier rebuttals and reduces the evaluators' fear of a supplier dispute, which, in turn, reduces rater dissonance and increases SPE efficacy. Relationship quality is degraded by evaluator turnover. Thus, once assigned, buying organizations should seek to retain supplier performance evaluators in that role. On complex, long-term contracts, evaluators, over time, develop a thorough understanding of the supplier's processes, deliverables, how to evaluate performance, and how to communicate and cooperate to accomplish the objectives of the contract. When this tacit knowledge is interrupted, relationships suffer. The bi-directionality and formality of communications also affect relationship quality. Thus, working level communication should not be stifled and formal channels to communicate should be established and maintained.

## Theoretical Implications

Perhaps most importantly, this research suggests that in order for SPEs to be effective in motivating higher levels of supplier performance and in mitigating the risk of future adverse selection, those consumers of the information (i.e., the current

suppliers and future buyers) must believe the SPE is true. As such, SPE accuracy and sufficient rating justifications become the essential factors explaining SPE efficacy. The acceptance of feedback affects employees' responses to feedback (Ilgen et al., 1979). "Acceptance refers to the recipient's belief that the feedback is an accurate portrayal of his or her performance" (Ilgen et al., 1979, p. 356). Further, constructs such as surveillance (i.e., monitoring) and leverage (i.e., opportunism) that are key components of agency theory and transaction costs economics are less impactful. Thus, of the four theories relied upon to explain SPE efficacy – agency theory, organizational behavior, social exchange theory, and channel communication, it appears that organizational behavior theory is paramount – specifically, human performance. This is extraordinary since most industrial buyer-supplier exchange literature relies primarily on theories such as agency theory, transaction cost economics, resource-based view of the firm, contingency theory, social exchange theory, channel communications, systems theory, and game theory. Of all theories, organizational theory represents only seven percent of theories relied upon in supply chain research (Defee et al., 2010). We now know which literature will likely be most instructive in pointing supply chain scholars to the most relevant phenomenon to explain supplier performance motivation and buyer uncertainty avoidance. Consistent with human performance literature, such phenomenon associated with the acceptance of performance feedback might include: (1) expertise of the source of feedback, (2) credibility of the source of feedback, (3) the recipient's trust in the source's motives of the feedback, (4) consistency of feedback, and (5) specific justifications for the feedback (Ilgen et al., 1979).

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 supplier is considered an agent of the buyer in promulgating the buyer's mission. Second, the buyer team is comprised of multiple agents to itself. In the case of multiple evaluators in different sub-organizations, 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 performance evaluators and supplier employees could begin to identify more with the program than with their employers. In other words, sometimes, what is advantageous for the program can supersede what is advantageous for either the buyer team or the supplier. This explains the halo effect afforded a supplier who fails in one instance of performance, yet the evaluator does not mention the failure in the SPE because of reluctance to taint the program or the supplier'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.

Additionally, this research identifies an omitted dimension of the economics of information theory, and thus expands its application from business-to-consumer contexts to business-to-business contexts. It often pays for a supplier to reduce a buyer's perceived risk of transacting with a supplier ex ante via use of costly signals such as: (1) advertising (to establish brand image and to make promises), (2) warranties (to signal high quality), and (3) premium prices (to signal high, unique value). However, as reported by suppliers, it also may pay to incur the transaction costs to conceal poor performance information and thereby preserve reputation - an intangible resource. If suppliers can dispute and effectively negotiate ratings, they have a chance to prevent future buyers from discovering their true reputation. This information cloaking serves as a sort of *reverse signaling* and can increase the buyer's ex ante costs of information acquisition. If contracts are not lost due to a buyer's ignorance, the transaction costs of concealment may be worth the effort.

This research discovered that there may be different forms of leverage employed by buy-side performance evaluators. *Proactive* leverage manifests as a threat, and is used to get a supplier to do something (e.g., improve performance). *Quid pro quo* leverage is a favor (e.g., a more favorable, inflated SPE) with an expectation of payback (i.e., a debt owed). Both forms of leverage are present in SPEs. More research is needed to understand whether the different forms have differing effects on SPE accuracy. Ethical decision making theory may explain why differences in types of leverage exist. There is significant variability among different individuals' ability to recognize ethical issues, and this recognition is a function of the individual's degree of ethical sensitivity (Sparks and Hunt 1998). In Rest's (1986) model, an individual identifies alternative courses of action and considers the likely consequences of each alternative as they affect the interests, welfare, or expectations of each party involved. Consequences of being exposed in exercising quid pro quo leverage may be perceived as more severe than those of proactive leverage.

Finally, this research contributes several new scales that reliably and validly measure key phenomenon in supplier performance management. A scale was developed to measure SPE efficacy, including its two components of motivating current supplier performance and mitigating the risk of adverse selection in a future source selection. This research also developed valid new scales for rating dissonance, rating justification, and fear of supplier dispute. These scales can be used in future research of buyer-supplier exchange.

## Study Limitations

This research is not without limitations. First, the sample size of the quantitative data is relatively small. A small data set precludes the use of more robust statistical techniques such as covariance-based structural equation modeling. Additionally, the response rate is quite low. A low response rate calls into question the external validity (i.e., generalizability) of the results and raises suspicion of systematic response biases. The response may have been subdued by several factors. First, the survey was necessarily lengthy. The fifty percent attrition rate suggests that many who started the survey did not finish it, and survey length could have been the cause. Another contributor to the attrition rate could be those officials who were not part of the target population who accessed the survey to validate its purpose. However, had all respondents who initiated the survey completed it, the response rate and numbers of respondents would still have been low. Second, past performance data is considered sensitive information. Even though measures were taken to mitigate this issue (e.g., an anonymous survey, no supplier identities collected, and no specific contract action identifiers collected), some prospective respondents may have been uncomfortable participating. A contributing factor may have been the recent data breach by the Office of Personnel Management involving the loss of sensitive information of 21 million government employees (Nelson and Tau, 2015). While the response rate is low, it is not uncommon in business research. Melnyk et al., (2012) revealed a sharp decline in response rates starting in 2002, with a steady decline of 1% annually. Five top journals reported low-end survey response rates ranging from 3% to 8%.

Another limitation of this paper is the lack of a quantitative test of emerged propositions surrounding the effects of SPE efficacy on supplier outcomes. Thus, while serving as a foundation, future research should expand and test the propositions developed herein. These propositions lend themselves well to cross-sectional data collected via survey. The research also employed a limited number of interviews. While rich insights were gleaned from experienced informants, other related phenomenon may be omitted with few informants.

SPE efficacy was measured by a six-item scale. However, only one of the six items measured the extent to which SPEs affect the supplier's motivation to increase performance. The other five items measured the extent that SPEs mitigate the risk of future adverse selection. Future research could refine the scale by adding more items to measure the former dimension of SPE efficacy thereby bolstering construct reliability and validity. Additionally, in the PLS SEM model, rating justification and rating dissonance accounted for only 31% of the variance in SPE efficacy. Future research could explore additional variables affecting SPE efficacy.

SPE efficacy was measured from the perspective of the buyer-side performance evaluator. However, the users of the information – future buyer teams who assess the risk of adverse selection and suppliers who inform whether the SPE motivated increased performance – would likely be better able to determine the level of SPE efficacy. Nonetheless, since evaluators monitor performance, they have insight as to whether suppliers attempted to increase performance. Additionally, many of the evaluator respondents served on the source selection team for the contract for which they reported or served on other source selection. Thus, the evaluators likely offer reliable judgement as to whether the SPE will be effective in mitigating future adverse selection.

Rating inflation and sufficiency of rating definition were each assessed using a single-item scale. For this reason, they were not modeled in the PLS SEM. Nonetheless, rating inflation and rating definition were analyzed in post hoc tests by forming groups of high and low values. These tests are contingent on the reliable and valid measure of high and low rating inflation and high and low sufficiency of rating definitions. Future research could develop scales to assess rating inflation and rating definition using multiple items thereby establishing the reliability of these latent constructs.

Relationship quality is comprised of three concepts per the measurement scale developed by Palmatier (2008) and adopted herein – trust, satisfaction, and commitment. In this sample, the three concepts did not all load on the same factor in the EFA; items measuring commitment loaded on a separate factor. This may be a nuance of a government sample since the Competition in Contracting Act limits long-term commitment.

The sample could be affected by self-selection bias. Those respondents to the survey (buyers) and interview informants (suppliers) who were highly dissatisfied with the CPARS policy and/or system, or perhaps highly satisfied and resistant to change, could have been more inclined to respond to the survey. Nonetheless, a review of the open comments fields from the survey showed a balance of favorable and unfavorable perspectives.

## Future Research Directions

Future research could quantitatively test the propositions surrounding the consequences of SPE efficacy and accuracy on suppliers. Such a comprehensive model with many variables and successive dependent variables could be tested via structural equation modeling.

One aspect of SPE efficacy concerns ongoing contractor performance management. 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 major systems acquisitions, the management of contractor performance on service contracts—where the Department of Defense spends the majority of its contracted funds—is often deficient and inconsistent (GAO, 2001). 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 its SPE system (i.e., CPARS) 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)? Future research could deploy a SPM system as a test case on a limited set of transactions. Using a quasi-experimental design, comparisons could be made to a control group that uses the organization's status quo means of supplier management.

# Conclusion

Organizations outsource a substantial portion of their missions. With increased reliance on suppliers, supplier performance management and risk reduction via supplier selection become paramount. However, organizations struggle to consistently, efficiently, and meaningfully evaluate supplier performance. This research examined the efficacy of supplier performance evaluations. Major factors affecting SPE efficacy include inaccurate evaluations and poor rating justifications. Consequently, often, performance information is not relied upon to make trade-offs in best value source selections. To explore the efficacy of supplier performance evaluations, this research first tested a conceptual model of key antecedents from the literature using a sample of performance evaluators of 131 contracts. Factors found to affect the efficacy of supplier performance evaluations include rating inflation, rating justification, and rating dissonance. Rating justification was affected by role overload, perceived accuracy and perceived usefulness. Perceived accuracy was affected by buyer fairness, leverage attitude, rating inflation, communication bi-directionality, communication formality, and the sufficiency of the requirement definition. Rating dissonance was affected by evaluator role overload and fear of supplier dispute, which was, in turn, affected by relationship quality and perceived accuracy. To explore consequences of SPE efficacy on suppliers, several suppliers were interviewed. The interview data was used to develop 36 propositions. From these findings, important managerial and theoretical implications are drawn and future research directions are identified. The central constructs involved in SPE efficacy appear to be perceived accuracy and rating justification. It is clear that this stream of research can pay significant dividends given the substantial reliance of organizations on suppliers.

THIS PAGE INTENTIONALLY LEFT BLANK

# References


Aberdeen Group. (2005, September). *The supplier performance measurement benchmark report*. Boston, MA: Author.

Armstrong, J. S. & Overton, T.S. (1977), "Estimating Nonresponse Bias in Mail Surveys," *Journal of Marketing Research*, 14 (3), 396-402.

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.

Berrios, R. (2006), "Government Contracts and Contractor Behavior," *Journal of Business Ethics*, 63, 119-130.

Blair, E. & Zinkhan, G.M. (2006), "Nonresponse and Generalizability in Academic Research," *Journal of the Academy of Marketing Science*, 34 (1), 4-7.

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 mangement* (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.

Carlsson, M. (2015). *Strategic sourcing and category management: Lessons learned at IKEA.* Kogan Page: London.

Churchill, G.A., Jr. (1979), "A Pardigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.

Cormican, K., & Cunningham, M. (2007). Supplier performance evaluation: Lessons from a large multinational organization. *Journal of Manufacturing Technology Management*, *18(*4), 352–366.

Council Of Defense And Space Industry Associations (2013), Ref: Navy Proposed Policy Letter: Superior Supplier Incentive Program, Memorandum dated May 7, 2013.

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.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.

Defee, C., Williams, B., Randall, W. and Thomas, R. (2010). An inventory of theory in logistics and SCM research," *International Journal of Logistics Management*, 21(3), 404-489.

Dillman, D. A. (2000), *Mail and Internet Surveys: The Tailored Design Method*, 2nd ed., John Wiley and Sons, Inc.: New York.

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.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2012).

Federal Acquisition Streamlining Act (FASA), (1994).

Fornell, C. and Larcker, D.F. (1981), "Evaluating Structural Equation Models With Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Ganster, D.C., Hennessey, H.W., & Luthans, F. (1983), "Social Desirability Response Effects: Three Alternative Models," *Academy of Management Journal*, 26 (2), 321-31.

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.

Government Accountability Office (GAO). (2014). Contractor performance: Actions taken to improve reporting of past performance information (GAO-14-707). Washington, DC: Author.

Griffis, S.E., Goldsby, T.J., and Cooper, M. (2003), "Web-based and Mail Surveys: A Comparison of Response, Data, and Cost," *Journal of Business Logistics*, 24 (2), 237-58.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis, 7th Ed.* Upper Saddle River, New Jersey: Prentice Hall.

Hair, J.F., Sarstedt, M., Ringle, C.M., and Mena, J.A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research, *Journal of the Academy of Marketing Science*, 40, 414-433.

Hair, J.F., Hult, G.T.M., Ringle, C.M., and Sarstedt, M. (2014). *A Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Los Angeles, CA: Sage.

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., Berkowitz, D. Muir, W., and Gravier, M. (2015). "Improving Services Supply Management In The Defense Sector: How the Procurement Process Affects B2B Service Quality", *Journal of Purchasing and Supply Management*, 21, 81-94.

Hawkins, T. (2013). Exploring the Efficacy of the Government's Current Use of Past Performance Information (NPS-CM-13-125). *Acquisition Research Program Sponsored Report Series*, Naval Postgraduate School.

Heide, J. B., & John, G. (1992). Do norms matter in marketing relationships? *Journal of Marketing*, *56*(4), 32–44.

Henseler, Ringle, & Sarstedt. (2015) A New Criterion for Assessing Discriminant Validity in variance-based structural equation modeling, *Journal of the Academy of Marketing Science*, 43(1), 115-135.

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.

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.

Johnson J, Sohi R, & Grewal R. (2004). The Role of Relational Knowledge Stores in Interfirm Partnering. *Journal Of Marketing*, 68(3), 21-36.

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

Kerlinger, Fred N. and Howard B. Lee (2000), *Foundations of Behavioral Research*, 4th Edition, Australia: Wadsworth Thomson Learning.

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.

Kline, R.B. (1997). *Principles and Practice of Structural Equation Modeling.* New York: Guilford Press.

Kumar,N, Scheer, L.K., & Steenkamp, J.E.M. (1995). The effects of supplier fairness on vulnerable resellers. *Journal of Marketing Research*, 32(1), 54-65.

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.

Landeros, R. & Plank, R.E. (1996), "How Ethical are Purchasing Management Professionals?," *Journal of Business Ethics*, 15 (7), 789-803.

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.

Melnyk, S. A., Page, T. J., Wu, S. J., & Burns, L. A. (2012). Would you mind completing this survey: Assessing the state of survey research in supply chain management. *Journal of Purchasing and Supply Management*, 18(1), 35–45.

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. (2011a). Supplier relationship management: An implementation framework. Tempe, AZ: CAPS Research.

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2011b). *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 relationshp 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.

Naval Seal Logistics Center Portsmouth (2014, July). Guidance for the contractor performance assessment reporting system (CPARS). Accessed online 27 Oct 15 at http://dai-global-developments.com/assets/cpars_guidance.pdf .

Naval Sea Logistics Center Portsmouth (2015). CPARS Metrics accessed 3 Nov 15 at https://www.cpars.gov/metrics.htm.

Nelson, C.M. and Tau, B. (2015). OPM Director Katherine Archuleta Resigns After Massive Personnel Data Breach. Wall Street Journal, retreived 8 October 2015 from http://www.wsj.com/articles/opm-director-katherine-archuleta-resigns-after-massive-personnel-data-hack-1436547193

Netemeyer, R.G., Boles, J.S., McKee, D.O., & McMurrian, R. (1997). An investigation into the antecedents of organizational citizenship behaviors in a personal selling context. *Journal of Marketing*, 61(3), 85–98.

Niskanen, W. A. (1968). The peculiar economics of bureaucracy. *American Economic Review, 58*(2), 293–305.

Nunnally, J.C. (1978), *Psychometric Theory*, 2nd ed., McGraw-Hill: New York.

Office of Management and Budget (OMB). (March 11, 2008). Update to Civilian Position Full Fringe Benefit Cost Factor, Federal Pay Raise Assumptions, and Inflation Factors used in OMB Circular No. A-76, "Performance of Commercial Activities". Memorandum M-08-13. Washington, D.C.

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.

Office of Federal Procurement Policy (OFPP) (2000, May). Best Practices For Collecting And Using Current And Past Performance Information. Washington, DC: Office of Management and Budget.

Office of Personnel Management (2014). *Common Characteristics of the Government Fiscal Year 2014*. Washington, DC: Office of Personnel Management.

Palmatier, R. W. (2008, July). Interfirm relational drivers of customer value. *Journal of Marketing*, *72*(4), 76–89.

Patrick Van Ecke, 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

Paulhus, D. (1991), "Measurement and Control of Response Bias," In J.P. Robinson, P.R. Shaver and L.S. Wrightsman (Ed.), *Measures of Personality and Social Psychological Attitudes: Volume I of Measures of Social Psychological Attitudes*, 2: 17-59, San Diego: Academic Press, Inc.

Perkins, W. S. (1993). Measuring customer satisfaction. *Industrial Marketing Management*, *22*(3), 247–254.

Podsakoff, Philip M. and Dennis W. Organ (1986), "Self-Reports in Organizational Research: Problems and Prospects," Journal of Management, 12 (4), 531-44.

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.

Randall, D.M & Fernandez, M.F. (1991), "The Social Desirability Response Bias In Ethics Research," *Journal of Business Ethics*, 10 (11), 805-17.

Rest, J.R. (1986), Moral Development: Advances in Research and Theory, New York, NY: Praeger.

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.

Sparks, J.R. and Hunt, S.D. (1998), "Marketing Researcher Ethical Sensitivity: Conceptualization, Measurement, and Exploratory Investigation," *Journal of Marketing*, 62 (2), 92-109.

Stump, R .L., & Heide, J.B. (1996). Controlling supplier opportunism in industrial relationships. *Journal of Marketing Research*, 33(4), 431–441.

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.

Undersecretary of Defense, Acquisition, Technology, and Logistics (2015). Implementation Directive For Better Buying Power 3.0 – Achieving Dominant Capabilities Through Technial Excellence and Innovation, Memorandum (April 9, 2015).

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.

Wieters, D. C., & Ostrom, L. L. (1979). Supplier evaluation as a new marketing tool. *Industrial Marketing Management*, *8*(2), 161–166.

Williamson, O.E. (2005). "The Economics of Governance." *American Economic Review*, 95(2), 1-18.

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.

THIS PAGE LEFT INTENTIONALLY BLANK

# Appendix A. Survey Invitation

Dear Sir or Ma'am,

You have been selected to participate in a study of supplier performance evaluation (SPE)/past performance. This research was approved by DASN RDA, and by the Office of Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-14), Navy Personnel Command, (deemed exempt from a report control symbol). The research is funded by a grant through the Naval Postgraduate School's Acquisition Research Program (N00244-15-1-0057).

The purpose of this study is to identify the factors affecting SPE efficacy, and how SPE efficacy, in turn, affects supplier outcomes such as performance. I respectfully request your assistance to complete the web-based survey located at the following link:

https://wku.co1.qualtrics.com/SE/?SID=SV_cAsofr69WYG75qd

Your participation is anonymous.

Your participation is voluntary. Responses are vital to conducting valid research that represents your knowledge and experience.

Please complete the survey no later than [date]. The survey should take approximately 25 minutes to complete.

For your time, you will be eligible to enter a random drawing for a new Apple iPad Mini, 16 GB, WiFi. To enter, follow the instructions at the end of the survey. (This raffle is not funded with federal grant funds.)

At the end of the survey, you will also have an opportunity to share ideas for improving SPE/past performance data collection and use.

If you have any questions about the study, please contact Dr. Tim Hawkins by email to timothy.hawkins@wku.edu or at 270.745.2412, or contact the WKU Institutional Review Board (IRB) Compliance Manager, Mr. Paul Mooney at 270.745.2129.

I know your time is valuable. Thank you so much for your support.

# Appendix B. Survey

You are invited to participate in a study of Past Performance Practices. Responses to this questionnaire will be used to analyze the government's use of contractor past performance information. **Your response is requested no later than [date]**.

This DoD-funded research is being conducted through Western Kentucky University. Participation from professionals, such as you, is very important for the success of this research. Your response will help the researchers analyze the government's collection and use of past performance information.

The questionnaire is **anonymous**; your responses cannot be linked to you. There are not necessarily "right answers."

Procedures. Your extent of participation in this research involves only the completion of this questionnaire.

Synopsis. This is both an anonymous and voluntary questionnaire. (Please note, in order to obtain consistent and usable results, it is important that you answer all questions). It will take most respondents approximately 30 minutes to complete the questionnaire.

Risks and Benefits. Your participation in this research poses no known risk. You will be asked questions pertaining to a contractor performance assessment report (CPAR). There will be no personal benefits beyond having contributed your expertise to this important research. Results of the survey will be used responsibly and protected against release to unauthorized persons. If desired, you may contact the researcher below if you would like to receive a report of the results of the study.

Confidentiality & Privacy Act. All records of this study will be kept confidential and, since responses are anonymous, your privacy will not be at risk. No information will be publicly accessible which could identify you as a participant. Responses will be maintained by WKU for three years, after which they will be destroyed.

Points of Contact. Should you have any questions or comments regarding this survey, please contact the Principal Investigator: Dr. Tim Hawkins, Lt Col (ret), USAF, 270-745-2412, timothy.hawkins@wku.edu. Any other questions or concerns may be addressed to the WKU IRB Compliance Manager, Mr. Paul Mooney, 270-745-2129.

**Thank you** for your time and your participation in this effort.

**By clicking on the "Proceed" button, I am acknowledging that I have read and understand this information, that I understand the nature and purpose of this study – including its risks and benefits, and that I agree to voluntarily participate in this online survey. I also understand that I may discontinue at any time simply by exiting this website.**

□ Proceed □ Exit

## Section 1

**Instructions**: Please answer the questions in this questionnaire pertaining to the selected CPAR (or ACCAS/CCASS for construction/A&E). Recommend you have a copy of the completed CPAR available as you complete the survey.

**Did you collaborate with at least one other person in completing this CPAR? In other words, the decision on what ratings would be assigned and the narrative content was not yours alone; you sought consensus with or input from at least one other person.**

□ Yes □ No

**What was the total number of people who measured, evaluated, reviewed and reported on this contractor's performance on this contract/task order/delivery order (even if done outside of the CPAR system)?**

_____

**What is the name of any automated system(s) (other than CPARS/CCASS) used to collect and/or track contractor performance information?**

_____

**For this contract/task order/delivery order, what type of CPAR was completed?**

□ Architect-Engineer (ACASS Module)

□ Construction (CCASS Module)

□ Systems

□ Non-Systems (Services, Information Technology, Operations Support, etc.)

**For this contract/task order/delivery order, what was the PSC/FSC code?** _____

**Please estimate the total number of hours spent by the Government team to complete the CPAR. Include only the time spent directly working on the CPAR (including collecting performance data), and exclude idle time awaiting action by another person. Include the time from all parties involved (e.g., performance evaluators and reviewers).**

_____

**On this contract, how many "out of cycle" CPARs have ever been completed?**

_____

**On this contract, how is the contractor's performance actively managed? (Note, "*actively manage*" herein means to continuously measure performance and to periodically communicate the buyer's assessment of performance to the**

**contractor to foster continuous improvements throughout the period of performance.) (Check all that apply):**

□ We engage the contractor only when problems arise
□ Monthly (or more frequently) performance reviews with the contractor
□ Quarterly performance reviews with the contractor
□ Semi-annual performance reviews with the contractor
□ Annual performance reviews with the contractor
□ Supplier scorecard
□ Rank-order contractors according to performance
□ Objective performance measures consistently used
□ Subjective performance measures consistently used
□ We don't manage contractor performance; that's their job
□ Other (Explain): _____

**On a scale of 1 to 7, where 1 represents "Completely Insufficient" and 7 represents "Completely Sufficient", please rate how sufficiently the CPAR ratings were justified.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following statements.**

**The rationale for the assigned CPAR rating was thoroughly documented.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**An inspector general would conclude that the CPAR rating was sufficiently explained.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**We did not have the factual data to support the rating that the contractor deserved.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**Anyone who reads this CPAR will understand the ratings based on the supporting information in the report.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**If I reported the contractor's performance accurately, the contractor would have disputed/rebutted the rating(s).**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**While completing the CPAR, at least one member of the government team was concerned that the contractor might dispute the assigned rating.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**The Government was lenient in the CPAR rating in order to avoid the conflict associated with the contractor's rebuttal.,**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**To report the ratings that the contractor actually deserved would have consumed too much time responding to the contractor's rebuttal.,**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**Someone on the government team either changed or influenced a change to the CPAR (at least one rating or narrative) in response to the contractor's rebuttal.,**

□ Yes □ No

**If you answered "Yes" to the previous question, do you agree with all of the changes?**

□ Yes □ No

**If the Reviewing Official (or any other individual) changed or influenced you to change at least one rating or narrative, why did he/she do so? (Check all that apply):**

□ N/A; Nobody changed or influenced a change to at least one rating or narrative.
□ To avoid time-consuming conflict
□ The factual data/justification did not support the Assessing Official's rating
□ The contractor's rebuttal had merit
□ To preserve the relationship
□ To protect the program
□ A lesser rating would tarnish the contractor's reputation
□ A lesser rating could hinder the contractor's ability to win future business
□ To gain bargaining leverage over the contractor
□ Other (Explain): _____

**On a scale of 1 to 7, where 1 represents "No Monitoring of the Contractor" and 7 represents "Extensive Monitoring of the Contractor", rate the <u>amount of government surveillance</u> of the contractor's performance in the following areas:**

a. Quality

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. Timeliness of Performance

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. Fulfillment of Performance Requirements in the specifications, drawings, Statement of Work, or Performance Work Statement

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. Compliance With Contract Terms & Conditions

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

The following section requests that you show the <u>final</u> ratings recorded in the CPAR.

**On the CPAR, were any rating categories not completed/rated?** □ Yes □ No

**How many ratings and narratives did the contractor disagree with (i.e., request to be changed) in the initial C** _____

**If a "<u>non-systems</u>" CPAR, what were the assigned ratings:**

**Quality of product/service**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Schedule**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Cost control (if applicable)**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Business relations**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Management of key personnel**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Utilization of small business**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**If a "systems" CPAR, what were the assigned ratings:**

**Technical (quality of product)**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Product performance**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Systems engineering**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Software engineering**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Logistics support/sustainment**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Product assurance**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Other technical performance**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Schedule**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Cost Control**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Management**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Management responsiveness**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Subcontract management**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Program/other management**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**Utilization of small business**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**If the contract was for construction, what were the assigned ratings:**

**Overall rating**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**If the contract was for architect-engineering, what were the assigned ratings:**

**Overall Rating**

□ Unsatisfactory □ Marginal □ Satisfactory □ Very Good □ Exceptional

**On the CPAR, how did you rate the following question: "Given what I know today about the contractor's ability to execute what he promised in his proposal, I _____ award to him today given that I had a choice."**

□ definitely would not
□ probably would not
□ might or might not
□ probably would
□ definitely would

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

**I am not given enough support to accomplish assigned objectives**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**It often seems like I have too much work for one person to do**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**The performance objectives on my job are too high**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**I am not given enough time to do what is expected of me on my job**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**It is ok for the Government to threaten the contractor with a lower CPAR rating.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**It is ok for the Government to use the CPAR as bargaining leverage with the contractor.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**If we give the contractor a CPAR that is better than what they deserve, the contractor should reciprocate in some way.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**Leverage can be gained by providing the contractor an overly favorable CPAR.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. The requirement was very well defined in the contract/task order/delivery order.
   □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
b. The contract/task order/delivery order (including the statement of work, performance work statement, specification, drawings, etc.) defined the requirement very well.
   □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
c. There were no flaws or omissions in the definition of the requirement (including the statement of work, performance work statement, specification, drawings, etc.).
   □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. There were no ambiguities in the definition of the requirement (including the statement of work, performance work statement, specification, drawings, etc.).
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. The requirement, as defined in the contract/task order/delivery order, expressed to the contractor exactly what we needed.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Completely Unfair" and 7 represents "Completely Fair", rate the following:**

1. To what extent was the contractor's performance fairly evaluated considering the contractual responsibilities?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
2. To what extent was the contractor's performance fairly evaluated given the amount of effort the contractor put forth?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
3. To what extent was the contractor's performance fairly evaluated given the challenges of the contract?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
4. To what extent was the contractor's performance fairly evaluated for the work the contractor did well?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7
5. The contractor deserved the performance ratings it received.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. The information in the CPAR was accurate.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. The record of the contractor's performance, as recorded in the CPAR, contains errors.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. The performance feedback in this CPAR is an accurate portrayal of the contractor's performance.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. The government accurately measured the contractor's performance level.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. The government consistently measured the contractor's performance level.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. This CPAR was based solely on factual performance information/data.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

g. All of the assessed ratings could be traced back to records of contractor performance.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

h. One or more of the final CPAR ratings was inflated (i.e., greater than the contractor deserved).
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

i. I use CPARs to actively manage the contractor's performance throughout the period of performance rather than solely to report the performance at the end of a period of performance. (Note, "*actively manage*" herein means to continuously measure performance and to periodically communicate the buyer's assessment of performance to the contractor to foster continuous improvements throughout the period of performance.)
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

j. Had ten other qualified people completed this CPAR, each would have arrived at the exact same ratings.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

k. Had ten other qualified people completed this CPAR, each would have arrived at the same justifications for each rating.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. Using the CPAR process/system improves my job performance
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. Using the CPAR process/system enhances my effectiveness on the job.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. The CPAR process/system helped me to report contractor performance quickly.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. Using the CPAR process/system makes it easier to do my job.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. The CPAR process/system ensured that I accurately report contractor performance.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. The CPAR process/system ensured that the contractor received a fair evaluation of performance.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

g. It is futile to report the real ratings that the contractor deserves since management will either change the ratings or make me change the ratings.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

h. If the FAR did not require me to complete a CPAR, I would not do it.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

g. Using CPARS in my job increases my productivity.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

h. I find CPARS useful in my job.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Very Infrequent" and 7 represents "Very Frequent", estimate the frequency of communicate with the contractor over a typical four-week period for each communication mode below:**

a. Face-to-face interaction
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. Telephone interaction
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. Written letters, correspondence

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. E-mail/Text Messaging
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. Web conference/VTC
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. Online chat
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "None" and 7 represents "A Lot":**

a. How much performance feedback do you provide the contractor?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. How much performance feedback does this contractor provide to you?
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**Rate the following on a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree".**

a. In coordinating our activities with this contractor, formal communication channels are followed (i.e., channels that are regularized, structured modes versus casual, informal, word-of-mouth modes).

b. We have a formal system to track the performance of the contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. We have a formal program for evaluating the contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. The source of our information about our evaluation program is predominantly word-of-mouth.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. Our evaluation process is conducted through standard procedures for collecting, analyzing, and reporting contractor performance information.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. Our organization regrets the decision to do business with this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. Overall, we are very satisfied with this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. We are very pleased with what this contractor does for us.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. Our organization is not completely happy with this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. If we had to do it all over again, we would still choose to use this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. This contractor keeps promises made to our organization.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. This contractor is always frank and truthful with us.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. We believe the information this contractor provides us.

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. This contractor is genuinely concerned that our organization succeeds.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. When making decisions, this contractor considers our welfare as well as their own.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. This contractor is trustworthy.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. We have a strong sense of loyalty to this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. We expect this contractor to be working with us a long time.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. We are really committed to developing a working relationship with this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. We see this relationship as a long-term alliance.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. This CPAR will help inform evaluators about this contractor's performance risk in a future source selection evaluation.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. If future source selection evaluators read this CPAR, they can assess the risk of dealing with this contractor.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. This CPAR will reduce future source selection evaluators' uncertainty about this firm's likelihood of performing similar work well.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. This CPAR can help future source selection evaluators to make a contract award decision.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. On this contract/delivery order/task order, I have seen/heard evidence that CPARS motivates the contractor to perform well.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. With this CPAR, future source selection evaluators can be confident in their assessment of the risk of this contractor successfully performing on a similar future contract.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On a scale of 1 to 7, where 1 represents "Strongly Disagree" and 7 represents "Strongly Agree", rate the following:**

a. Between myself, the "Reviewing Official," and other performance evaluators, there was some disagreement on at least one CPAR rating.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

b. Significant effort was required to deliberate with others as to what rating(s) to assign.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

c. At least one performance evaluator and/or "Reviewing Official" believed that at least one rating or narrative justification should have been different.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

d. The government team had difficulty reaching consensus on the ratings or narrative justification.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

e. Each member of the government team evaluating performance completely agreed with each assigned rating.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

f. While completing the CPAR, I was concerned about whether someone else would disagree with the ratings.
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**If there was disagreement within the government team on any aspect of the CPAR narratives or ratings, please explain why.**

_____

**Before the CPAR was entered into the CPAR system, the Government solicited input from the contractor about its view of what ratings or narrative justifications should be in the CPAR.**

□ Yes □ No

**Before the CPAR was entered into the CPAR system, the contractor submitted input about its view of what ratings or narrative justifications should be in the CPAR.**

□ Yes □ No

**If yes, on a scale of 1 to 7 where 1 represents "None" and 7 represents "A Lot", to what extent did the final CPAR ratings and narrative justifications resemble the contractor's input?**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**On this contract, how many relationship connections existed between individual government employees and contractor employees? Note, in cases where one contractor connected with several government employees (and vice versa), count each as a seperate connection.** _____

**Note: The following questions are general in nature. They do not necessarily pertain to the chosen CPAR.**

For the most recent source selection in which you were involved and that included past performance as an evaluation factor for award, to what extent did the past performance rating affect the source selection/award decision, on a scale of 1 to 7
□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □ N/A

**What type of incentive does this contract/task order/delivery order contain – if any? (Check all that apply):**

□ Award Fee □ Cost-Based Incentive Fee □ Award Term □ Performance-based Payments

□ Performance-Based Incentive Fee □ Delivery-Based Incentive Fee □ Liquidated Damages clause

□ Other (Please Explain): _____

□ This Contract Does Not Contain an Incentive

**What is the contract type?**

□ Fixed Price □ Cost Reimbursement □ Time and Materials □ Labor-Hour □ Other

□ Hybrid (multiple contract types)

**What type of supply or service is the contractor providing?**

□ Service □ Construction □ Supplies/Commodities/Spares □ Weapon System

□ Other Capital Equipment

**Program/contract description** _____

**What is the current total dollar value of the contract/task order/delivery order (including all options)?**

_____

**At the time you completed the CPAR, what was the total duration of the contract/task order/delivery order (in months) from the date of contract award?** _____

**Was this contract/task order/delivery order competed?** _____

**On a scale of 1 to 7, where 1 represents "Not at all Critical" and 7 represents "Extremely Critical", rate the criticality of the contracted item or service to your agency's mission.**

□ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7

**The contractor for this past performance evaluation is a:**

□ Small business (includes SB, SDB, 8(a), woman-owned SB, veteran-owned SB, & HUBZONE)

□ Large Business

□ Ability One, Federal Prison Industries

**By which agency are you employed?** _____

**Were you involved in the source selection (or negotiation if sole source) for this contract/task order/delivery order?** □ Yes □ No

**What is the highest level of Acquisition Professional Development Program (APDP) certification that you hold?**

□ APDP Level 1 □ APDP Level 2 □ APDP Level 3 □ No APDP Certification

**What is your career field?**

□ Quality Assurance □ Program Management □ Contracting □ Engineering □ Logistics □ Other

**I am a:**

□ Program Manager □ Contracting Officer/Specialist □ Quality Assurance Evaluator/COR/Inspector/COTR

□ Product/Service End User □ Other (Explain): _____

**The "Reviewing Official" was a: (Check all that apply)**

□ Program Manager □ Contracting functional □ Product/Service End User □ General Officer/SES

□ O-6/GS-15 □ Other _____

**Over the life of this contract, how many times has a performance evaluator changed/turned over?** _____

**In how many different physical locations did contractor performance occur?**

_____

**Were you involved in writing or determining the technical specifications/statement of work for this contract/task order/delivery order?**

□ Yes □ No

**What is the highest level of education that you have attained?**

□ High School Diploma / GED □ Associates □ Bachelors □ Masters □ Doctoral / Professional

**How many years of experience do you have evaluating contractor performance?**

_____

**What is your gender?**

□ Male □ Female

**Can you think of any policy changes needed to improve the government's collection and/or use of past performance information? (optional)**

_____

**We appreciate any comments or feedback you can provide on the topic of past performance in government contracting and/or this survey. (optional)**

_____

# Appendix C. Interview Questionnaire

Are CPARs important to your company?

- If so, why?

Does your company have any dedicated positions/personnel whose job it is to manage/oversee CPARs?

How does the government's CPAR process and IT system differ from those of your for-profit-sector clients?

- Do for-profit-sector clients tend to give lower or higher (or the same) supplier ratings?
- Do for-profit-sector clients offer a chance to respond to ratings?

Are the past performance reports value-added? Why or why not?

What are the greatest weaknesses associated with the CPAR process and IT system?

- Why are these weaknesses important?
- How would you "fix" any CPARS processes and/or IT system if you could?

The purposes of the government's past performance information collection and use policy is to: (1) inform source selection teams to mitigate the risk of adverse selection and (2) to motivate contractors to perform well. Do you believe the CPARS processes and IT system accomplish these objectives? Why or why not?

- Do past performance evaluations motivate you to alter your effort/level of performance?
- Do you believe past performance evaluations are effective in reducing a source selection teams' uncertainty with respect to future performance of a prospective contractor?

Have you ever received any past performance evaluations (ratings and/or narratives) that were less favorable than you expected?

- If so, how many times?
- Explain/elaborate.
- What was the source of the disconnect in understanding?

What communication strategies do you employ, if any, to mitigate surprises?

Have you observed any issues with the timeliness of assessments (i.e., time from the end of a period of performance to receipt of CPAR)?

Do you ever discuss CPARS ratings with a buyer representative prior to receiving the assessment in the CPARS system?

What kinds of differences in CPAR scoring methodology do you experience across multiple contracts and/or multiple government clients?

Have you ever suspected or witnessed different perceptions of your firm's performance from different buyer-side evaluators or their managers/leaders on the same contract action?

- If so, why might different people on the buyer team hold different perceptions of your performance?
- How do you react to or manage these difference?

Have you ever received an inaccurate, incomplete, or misleading past performance evaluation?

- If so, why might a past performance evaluation be inaccurate, incomplete, or misleading?

Have you ever received a past performance evaluation with an inadequate justification for ratings?

How important is your company's reputation?

Is it your philosophy (or is it your company's philosophy) to contest any/all negative ratings and/or narratives?

Do you or your company hold the position that you should protect the company's reputation at all costs?

- Has your company sought a rating change even though the rating was warranted?
- Are the man-hours required to get a rating changed worth it?
- Does your urgency to avoid negative evaluations change with the competitiveness of the market in which you do business?
- Is the accuracy of the CPAR less important than your company's reputation?

Have you ever, or would you ever, offer concessions in order to alter/improve an anticipated negative CPAR?

Has a less-than-expected CPAR rating ever caused your company/employees to change anything with how they perform the contracted work (i.e., "get even" informally)?

To what extent does the government appear to use the past performance evaluation as a way to actively manage performance throughout the performance period?

Do poorly defined requirements or differences in expectations play a role in your satisfaction with a past performance evaluation?

Why do customers change ratings?

- Do customers ever trade a more favorable rating for some concession from the contractor?

How have you reacted to past performance evaluations that appear inaccurate?

- How would that affect the relationship with the customer?
- How would that affect how much you value or prioritize the customer?
- What factors do you consider when deciding whether to rebut a PP rating?

How does buyer team turnover affect past performance evaluations?

On average, how many man-hours do you spend responding to a PP evaluation (from receipt to completion of a report)?

Do you ever suspect buyer opportunism in ratings?

- Do you suspect that the government ever uses the PP rating/evaluation as leverage?

Is the past performance evaluation process fair?

Do you see inconsistent definitions of the ratings across CPARS?

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

**Demographics:**

How many past performance evaluations have you personally participated in?

_____

How many past performance evaluations has your organization participated in?

_____

Duty title: _____

Years of experience managing customer evaluations? ____

Gender: _____

Industry: _____

Compete in competitive market(s)? _____

Business size: Large or Small?

# Appendix D. Survey Respondent Recommendations (Assessors)

| Can you think of any policy changes needed to improve the government's collection and/or use of past performance? | Code |
|---|---|
| • "CPARS should not have a cost factor or it should be written differently since most of the time the cost are based on government's budget and the Contractor cannot impact cost except with personnel." | O |
| • "Consistent application of rating criteria across DoD. I have had contractors ask for their rating to be increased to bring it in line with what other services provide for the same service or product form the safe facility." | C |
| • "1) Require AOs to complete and submit their assessments with 30 days of end of the period of performance. / 2) Implement independent or outside performance evaluation of activity/organization Focal Points and Assessing Officials during their own individual performance assessments and retrain those coming up short." | A |
| • "Not make it mandatory for source selection" | O |
| • "Under the new CPARS policy, the new evaluation area of "REGULATORY COMPLIANCE" was added. This new evaluation area/criteria should be either "complied/Satisfactory" or "non-complied/Unsatisfactory" which is the true/accurate assessment for this area because this area simply can't be anything else (i.e., Exceptional, Very Good, Marginal)." | CD |
| • "The Past Performance Evaluation is complex yet depending on which lawyer or contracting officer you have their approach to this factor is very different. More training on the application of the Past Performance within source selection is important. Additionally this act of evaluation for CPARS is a convoluted process and depending upon demeanor of an individual there could be very different ratings." | T<br>A |
| • "Allow greater accuracy of reporting." | A |
| • "Service contracts are difficult to evaluate with the current CPARS measures" | CF |
| • "CPARS did not assist at all in the evaluating past performance for future contracts. Most CPARS I reviewed had no narrative, so were not used." | I |
| • "I think there should be more formal training for CPAR writers. There was no syllabus or anything for how to write CPARs, and I initially had to write CPARS on about 6 contracts that were complete before I checked in. I read the guidance and solicited | T |

| | |
|---|---|
| input from the people who actually worked on the contract and my appraisals went through a reviewing official who was familiar with the contracts, but I felt odd writing appraisals on contracts that I didn't work on." | |
| • "It's not agency policy but proper tracking of CPARS being completed. Additionally, each Command should look had which CPARS rating were assigned. If over 50% of the ratings are Outstanding or Very Good, someone needs to review the CPARS to make sure the justification support the rating." | C |
| • "An interim CPARS, though labor intensive, would improve the overall process." | T |
| • "Standardization across platform types. Limited CPARs requirements for Congressionally directed contractors." | C |
| • "Collection of data must be balanced with the cost of having it reported to the government, and then analyzed across the distributed receiving team." | TC |
| • "Would be more helpful if collected more often." | F |
| • "Past Performance is a farce. We are not allowed to use Past Performance as a negative criteria [sic]. There is [sic] at least four companies that I know of where their Past performance is less than optimal and they are arrogant about it. There is one company that tells the government, "To bad, this is what we are giving you"." | O |
| • "Consistent application of ratings." | C |
| • "It is important that the rationale support the rating. Any rating higher than satisfactory needs a really supportive rationale." | J |
| • "I would like PP that is relevant to natural resources. Also, CPARs does not consider that everything depends on the subcontractor." | S |
| • "The CPAR does give the contractor a sense of how things are going from the Government's perspective, but I have NEVER known of a CPAR evaluation actually being used in a source selection." | PU |
| • "Evaluators do need to follow the CPAR guidance and write objective narratives. Ratings narratives that come across as "they're a bunch of nice guys" do nothing for a source selection team, and I have seen quite a few situations like I mention. You then end up calling the authors to find out what the contractor's real performance is." | J<br>A |
| • "Things work very well as long as the correct ratings are assigned and not over inflated." | A |
| • "The contracting entity should be required to review the evaluation in the system prior to release to the contractor. Our organization has determined that a contracting officer be the Reviewing Official as many previous reports were forwarded to | O |

| | |
|---|---|
| the contractor with information that did not match the documentation in file (inspection results) and/or provided feedback on issues not directly related to contractor performance" | |
| • "The government needs contractors to be better estimators of expected cost and technical performance. Program managers must be able to trust the information that is provided by the contractor and have reasonable assurances that the numbers are realistic and not provided to "win the contract" or otherwise look favorable with respect to competitors - only to later fail to deliver because there are no consequences for providing 'bad estimates'. Therfore [sic], contracts should incentivize adherence to cost estimates. Additionally, the contractors track record of accurately estimating work should be available to the govt to assess risk of a proposal for a solicitation." | O |
| • "Focus on contractors who fail to perform. Identify contractors who have reinvented themselves under a new company name in order to resurface and bid for contracts." | O |
| • "CPARS is a valuable tool for assessing the contractor's performance, however when using previous CPARS during source selection for a new contract there is no value added on recurring service related contracts. A CPAR from another location, Good or Bad, is a direct reflection of the individuals at that location and in most instances the company the is awarded the contract hires the in-place employees." | PU |
| • "We could incorporate the PARs and CARS generated during the reporting period as backup or source data accessable [sic] to the COR, COR supervisor, PCO, Assessing official and reviewing official" | J |
| • "Consistency on the evaluations (Government wide) seems to be our major concern." | C |
| • "Past performance training for all involved in the input and rating of contractors' performance." | T |
| • "currently in source selection, the past performance of an offerror is boiled down to a yes/no - if they did have past performance, and it was satisfactory or better, then their past performance was rated acceptable, otherwise not acceptable. This is not enough information... there should be a larger range that the rating can fall within. The current methodology renders the past performance teams efforts meaningless." | PU<br>CD |
| • "Keeping track of emails, phone conversations, or site visits would help the government to generate the CPAR." | J |
| • "I think that the CPAR database needs to be accurately and constantly populated and CPARS completed. I've found that during Source Selection Evaluations that not all contractors had a CPAR avaliable [sic] for past projects although a DoD project | F<br>C<br>A |

| | |
|---|---|
| was presented for past performance review. I think the CCASS ratings are too subjective." | |
| • "There also seems to be some discretion on CPAR evaluation working from Purple for basic performance of the contract. Other start at a Green rating for basic performance and the contractor has to do over and above the contract to recieve [sic] a purple rating. Disparity of starting points" | C<br>CD |
| • "A better mechanism to align and compare the technical aspects of the work performed to the SOW/RFP being evaluated in a source selection." | RD |
| • "Consistent application of the standards needs to be applied across program offices." | C |

| **We appreciate any comments or feedback you can provide on the topic of past performance in government contracting.** | **Code** |
|---|---|
| • "In order for CPARS to truly be effective, major Commands need to look at the ratings for the CPARS for all contracts they have. [A particular] command did this and additional training has resulted. This kind of oversight helps foster more honest and accurate ratings, than just "filling" in the box." | MO |
| • "It would be great to have access to the Primary contractor performance rating od [sic] all sub-contractors." | S |
| • "Need to improve government's awareness of why CPARs exist - that is, to inform future government source selection and protect the taxpayers." | PU |
| • "Services need separate CPARS rating system - our PWS requirement is 100% performance, so no "exceptional" rating can ever be justified in the current system but several of our contractors perform such that an "exceptional" rating would be useful as a discriminator in source selections." | CF<br>CD |
| • "The close out CPAR should be a collective of the entire performance of the contract." | O |
| • "There were times that the contracting officer would request we change the CPARS and did not want to give a negative review. Not specifically this contract but other that I was a COR on." | A |
| • "CPARs DEFINITELY made a difference in the Contractors [sic] performance following low marks. It was the ONLY way we could get the message across. They felt they were so large they were immune to a low CPAR." | PU |
| • "Should be tailored "down" for smaller efforts." | O |
| • "The evaluation criteria [sic] is such that if a contractor completes | CD |

| | |
|---|---|
| the assigned task, no matter how much pain and suffering the contract specialist, project manager, quality assurance specialist, technical representatives, and others, must go through to facilitate the contractor's completion, the rating is Satisfactory. The ratings do not accurately portray contractor performance, and actual performance needs to be hidden in the narratives. This needs to change." | |
| • "First step is getting the Government to use CPARS, and then make sure the justification support [sic] the rating. One way our senior contracting officer does this is having a monthly COR report which requires a write up on the same categories as a CPAR. This way you can use the monthly reports to help do the CPARS rating." | PU<br>J |
| • "Collection/reporting of PP is extremely valuable for future awards, yet because it is so subjective (moreso [sic] in the services arena than in production), it is not 100% reliable, and negative performance is often explained-away by the vendor. Often it is iimpossible [sic] to contact the original evaluator due to personnel turnover or simple disinterest." | C<br>A<br>TU |
| • "the past performances of contractors is only as accurate as the evaluation person's ability to assess the task completions and deliverables to conform to tasking specifications. I [sic] technically qualified individual should evaluate technical products..." | O |
| • "The entire Procurement & Source Selection system needs to be reassessed, because if everyone were following regulation there would be a few powerful companies that should not be allowed to perform government work. not all contracts fit the same evaluation criteria." | A<br>CD |
| • "Perhaps allow the uploading of support documentation, i.e., cure notices, exceptional performance, etc." | RD |
| • "The categories for the ratings are not always the most appropriate for assessing the contractors [sic] performance. Fro [sic] example, a major portion of this contract is performing maintenance on an aircraft, however the only way to assess the performance is under quality, which does not capture the true nature of the performance. There are so many other facets to maintenance such as supplying parts on time, recovery of off-site aircraft, etc." | CD |
| • "I'd like to see the PWS of the contract rated attached in the CPAR system. This is extremely valuable to determine if the contractor really has done a similar effort of like magnitude." | RD |
| • "CPARS ratings need to be accurate. I served on a past performance evaluation team and one of my observations was that a vast majority of CPARS I reviewed were very over inflated. They would indicate "Exceptional" performance, but there was no narrative or justification in many cases." | A |
| • "we are wasting time evaluating every contract when we should be focusing on those who fail to perform. Those who fail to perform | O |

| | |
|---|---|
| should be documented in CPAR. The contract performance should simply be considered acceptable for the contracts with no documented performance problems." | |
| • "It's a good tool and one of the few effective levers I had on my contractor." | O |
| • "Collecting contractor performance feedback from CORs at the task order level is very time comsuming [sic]." | O |
| • "I believe if we want to enhance the process we should institutionalize the mid-term (or any) feedback. I believe our Program Manager are only using the CPAR because [sic] it is mandated and not as a Program Management tool." | F |
| • "I have observed many less effective uses and implementation of CPARs over my career. I have experienced situations where leadership has asked/directed me to change ratings etc. Those were difficult situations and in the end I did not change my rating. Often times, CPARs are to [sic] generic to help in source selection. Not everyone puts in the time to write an effective/accurate assessment." | A |

Key

Code, description, (frequency count)
O=other (12)
A = accuracy (11)
C = consistency (9)
CD = criterion/rating definition (7)
PU = perceived usefulness questioned (6)
T = training (5)
J = rating justification (5)
F = more frequent reporting (3)
RD = requirements document (3)
CF = criteria fit (with type of work) (2)
S = subcontractor PP omitted (2)
I = incomplete (1)
TC = transaction costs (1)
MO = management oversight (1)
TU = turnover (1)

# Appendix E. Dissonance Reasons

| |
|---|
| “Some program managers had difference in opinion with assessor.” |
| “Debated if "Technical" should be the average of the sub groups (product performance, system engineering, etc.) or the lowest of the sub groups. Also debated if requirements were exceeded by "some" or "many" to the Government's benefit.” |
| “Ratings were given by several different government COR some rated harder then [sic] others so that had to be neutralized.” |
| “Reviewers tend to be more concerned with how the contractor will react (in the case of a negative mark/report) than standing firm with facts to protect the taxpayer and ensure fairness in future evaluations.” |
| “A misinterpretation on the teams part in relation to color. They were initially giving blue ratings for meeting contract requirements. This was corrected to green unless it was shown that the contractor was meeting the definition of blue performance.” |
| “Different interpretations as to what the (negative or positive) impact of actions are by the contractor and how this should be assessed in the CPAR.” |
| “The rating was higher than the written justification.” |
| “Individual members of the team, have different performance expectations and interpetations [sic] of the ratings.” |
| “Different levels of interaction with the contractor resulted in different points of view regarding the performance of the contractor.” |
| “Some personnel involved with the review of the CPAR are not actively involved in the contract as much.” |
| “Although ratings were entered based on performance, reviewers requested changes based on assumptions of negative feedback from the contractor, or that although performance issues were identified, there was no net impact on cost or schedule.” |
| “The COR did not provide enough back-up information to justify a rating of marginal, therefore it was increased to satisfactory.” |
| “Reviewer thought some grades were too lenient without sufficient evidence to justify” |
| “Disagreement was avoided by not being as critical in the ratings in some areas.” |
| “Different points of view on the same issue: technically acceptable at twice the planned cost.” |
| “The challenge was arriving at satisfactory wording to describe the factors during the period of evaluation.” |
| “people have different perceptions of reality.” |
| “Any disagreement was due to the perspective of what position they held and an understanding of what was being evaluated.” |
| “Management tried to rate the contractor against performance that was not a contractual requirement.” |
| “Mis-understanding of the CPAR rating system; lack of documented eveidence [sic] to challenge rating (e.g. somewhat personal preference/emotional response).” |

| |
|---|
| “The CORs and the Contracting Officer sometime view the worked performance by the contractor differently. The Contracting Officer is supposed be objective and able to look at both sides to make sure the Government is getting what they paid for and to also to insure the contractor is treat [sic] fairly. The CORs sometimes only see if from their perspective.” |
| “The PWS was not as defined as some groups would of liked it” |
| “The way the performance metrics are laid out, the contractor cannot perform exceptionally. Only poorly or satisfactorily. This drives my ratings to be "satisfactory", where, as the COTR, I think the contractor did better than that in some areas.” |
| “We had a couple evaluations (Gov't & Contractor) that had different perspectives on the work performed. The Gov't evalutors maked [sic] down the CPAR becuase [sic] there was a problem. The Contractor self identified the problem and wanted credit for that self identification and resolution. There was room for interpretation within the rating & contract which caused a disagreement. The disagreements had to be resolved via CPAR assessing official.” |
| “Disagreement over Very Good vs Exceptional, as in the benefit to the government” |
| “Some people would consider work by this contractor on one of our other contract with them.” |
| “There were development and sustainment contracts with the same contractor and some of the evaluators were conflating the two, separate, contracts.” |


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

www.acquisitionresearch.net