



**Calhoun: The NPS Institutional Archive**  
**DSpace Repository**

---

Acquisition Research Program

Faculty and Researchers' Publications

---

2016-05-01

# Antecedents And Consequences Of Supplier Performance Evaluation Efficacy

Hawkins, Tim; Gravier, Michael

Monterey, California. Naval Postgraduate School

---

<http://hdl.handle.net/10945/53499>

---

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

*Downloaded from NPS Archive: Calhoun*

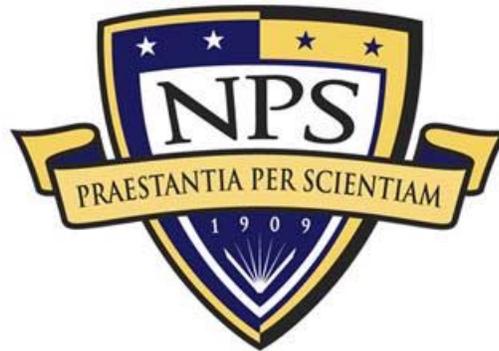


Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

**Dudley Knox Library / Naval Postgraduate School**  
**411 Dyer Road / 1 University Circle**  
**Monterey, California USA 93943**

<http://www.nps.edu/library>

SYM-AM-16-059



# PROCEEDINGS OF THE THIRTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

---

## THURSDAY SESSIONS VOLUME II

### **Antecedents and Consequences of Supplier Performance Evaluation Efficacy**

Timothy Hawkins, Lt Col, USAF (Ret.), Assistant Professor, Western Kentucky  
University

Michael Gravier, Associate Professor, Bryant University

**Published April 30, 2016**

Approved for public release; distribution is unlimited.

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



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](http://www.acquisitionresearch.net)).



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

# Panel 15. Connecting Contracting Strategy to Acquisition Outcomes

---

Thursday, May 5, 2016	
11:15 a.m. – 12:45 p.m.	<p><b>Chair: Major General Kirk F. Vollmecke, U.S. Army, Program Executive Officer, PEO IEW&amp;S</b></p> <p><b><i>Contract Design, Supply Chain Complexity, and Accountability in Federal Contracts</i></b></p> <p>Adam Eckerd, Assistant Professor, University of Tennessee Amanda Girth, Assistant Professor, The Ohio State University</p> <p><b><i>An Approach for Modeling Supplier Resilience</i></b></p> <p>Kash Barker, Associate Professor, University of Oklahoma Jose E. Ramirez-Marquez, Associate Professor, Stevens Institute of Technology Seyedmohsen Hosseini, PhD Candidate, School of Industrial and Systems Engineering, University of Oklahoma</p> <p><b><i>Antecedents and Consequences of Supplier Performance Evaluation Efficacy</i></b></p> <p>Timothy Hawkins, Lt Col, USAF (Ret.), Assistant Professor, Western Kentucky University Michael Gravier, Associate Professor, Bryant University</p>



# Antecedents and Consequences of Supplier Performance Evaluation Efficacy

**Timothy G. Hawkins,<sup>1</sup> Lt Col, USAF (Ret.)**—is an Assistant Professor in the Department of Marketing at Western Kentucky University in Bowling Green, KY. 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 Purchasing and Supply Management*, *Journal of Business Research*, *International Journal of Logistics Management*, *Industrial Marketing Management*, *Journal of Business Ethics*, *Supply Chain Management: An International Journal*, *Journal of Marketing Channels*, *Air Force Journal of Logistics*, *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 performance-based logistics, electronic reverse auctions, procurement ethics, buyer–supplier relationships, strategic sourcing, services procurement, and supplier performance management. Prior to his academic career, Dr. Hawkins worked in Global Procurement for NCR Corporation and served as a warranted contracting officer for the U.S. Air Force. [timothy.hawkins@wku.edu]

**Michael J. Gravier**—is an Associate Professor of Marketing and Global Supply Chain Management at Bryant University. He received his PhD in Marketing and Logistics from the University of North Texas, an MS in Logistics Management (with a specialization in transportation management) from the Air Force Institute of Technology, and BA with majors in Spanish and Anthropology from Washington University in St. Louis. Prior to his academic career, he spent 12 years as an active duty logistics readiness officer in the U.S. Air Force and worked as a research assistant in the Washington University School of Medicine. He has research interests in supply chain ethics, procurement, logistics pedagogy, logistics in emerging countries, and the evolution of supply chain networks in response to risk factors like obsolescence and changing information needs. [mgravier@bryant.edu]

## Abstract

There are numerous weaknesses associated with industrial buyers' collection and use of supplier performance evaluation (SPE) information (a.k.a., past performance information). 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 will employ a mixed method of qualitative interviews and quantitative analysis of survey data collected from suppliers and from assessors of supplier performance.

## 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

---

<sup>1</sup> Corresponding author



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 & Ostrom, 1979), and now SPE is an essential best practice in business-to-business sourcing (Gordon, 2008; Talluri & 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 & 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 & Ellegaard, 2011; Buffa & Ross, 2011), which invites different perspectives of supplier performance. To what extent does 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-based 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 suppliers. This research will explore 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 will be 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 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. First, a literature review is presented describing the conceptual framework and hypotheses. Next, the study presents the research designs and methodologies.



## Literature Review

### **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. Necessarily, then, agencies must collect and report contractor past performance information from government contracts (FAR Part 42) surpassing certain dollar values.

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 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. Suppliers deemed a superior supplier are eligible for relaxed or more favorable contract terms and conditions. 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 administrative burdens.

### **Supplier Performance Evaluation**

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

SPM pays off; a study by the Aberdeen Group (2005) found that supplier performance of companies with an SPM system improved significantly more than did the supplier performance of firms with no SPM system. Specifically, firms using an SPM system realized 10% greater price savings, 12% better on-time delivery improvement, four times greater quality improvement, and a 4% greater improvement in service. One large telecommunications firm realized a 290% reduction in the number of suppliers and a 260%



reduction in the value of inventory held due to an SPM system (Cormican & Cunningham, 2007). Another study (Limberakis, 2011) found that “best-in-class” buyers (1) are much more likely to benchmark supplier performance against others in the same industry, (2) achieved substantially higher percent on-time delivery (88% versus 48% for “laggards”), and (3) transacted with suppliers that experienced fewer catastrophic failure (2% versus 5% for other buyers). Of the best-in-class buyers, 63% had a supplier benchmarking and performance monitoring information technology system in place. 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. 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 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 research has explored why buyers decide to change their evaluations.

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). 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. 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 criteria 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. 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 in which 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., 2015), dissonance in supplier performance ratings should be commonplace. Buffa and Ross (2011) identified evaluator turnover as having a potential impact on supplier evaluations over time. Therefore, it is posited that

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 of the SPE.

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

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.

### **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) 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 & 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 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 a SPE context, fairness pertains to the extent to which the supplier is given the performance ratings it deserves. 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.

### **Methodology**

This research will employ quantitative and qualitative methodologies to examine the antecedents and consequences of supplier performance evaluation efficacy (Table 1). First, the quantitative methodology and results are detailed, 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

The quantitative method will examine data collected via survey of the personnel with the requisite knowledge of contractor performance, CPARS assessing officials. The hypotheses will be 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.

The qualitative method entails collecting data via interviews with suppliers whose performance has been rated. 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.

## 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.

## References

- Aberdeen Group. (2005, September). *The supplier performance measurement benchmark report*. Boston, MA: Author.
- 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.
- 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.
- Campbell, D. J., & Lee, C. (1988). Self-appraisal in performance evaluations: Development versus evaluation. *Academy of Management Review*, 13(2), 302–314.
- 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.
- Cormican, K., & Cunningham, M. (2007). Supplier performance evaluation: Lessons from a large multinational organization. *Journal of Manufacturing Technology Management*, 18(4), 352–366.
- Cousins, P. D., Lawson, B., & Squire, B. (2008). Performance measurement in strategic buyer–supplier relationships: The mediating role of socialization mechanisms. *International Journal of Operations & Production Management*, 28(3), 238–258.
- Dowst, S. (1972, June). How designers rate suppliers—and purchasing. *Purchasing*, 12, 87–93.
- Dwyer, F. R., Schurr, P. H., & Oh, S. (1987, April). Developing buyer–seller relationships. *Journal of Marketing*, 51, 11–27.
- Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2012).
- Federal Acquisition Streamlining Act (FASA). (1994).
- Gaski, J. F. (1984). The theory of power and conflict in channels of distribution. *Journal of Marketing*, 48(3), 9–29.
- Gordon, S. R. (2008). *Supplier evaluation and performance excellence: A guide to meaningful metrics and successful results*. Ft. Lauderdale, FL: J. Ross.



- GAO. (2009). *Federal contractors: Better performance information needed to support agency contract award decisions* (GAO-09-1032). Washington, DC: Author.
- GAO. (2014). *Contractor performance: Actions taken to improve reporting of past performance information* (GAO-14-707). Washington, DC: Author.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & 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., & 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). Monterey, CA: Acquisition Research Program, Naval Postgraduate School.
- Heide, J. B., & John, G. (1992). Do norms matter in marketing relationships? *Journal of Marketing*, 56(4), 32–44.
- House, R., & Rizzo, J. (1972). Role conflict and ambiguity as critical variables in a model of organizational behavior. *Organizational Behavior and Human Performance*, 7(3), 467–505.
- Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Performance*, 36, 16–78.
- John, G. (1984). An empirical investigation of some antecedents of opportunism in a marketing channel. *Journal of Marketing Research*, 21(3), 278–289.
- Kelman, S. (2010). Three ways to revitalize the use of past-performance data. *Federal Computer Week*. Retrieved from <http://fcw.com/blogs/lectern/2010/05/steve-kelman-past-performance-source-selection.aspx>
- Kern, D., Moser, R., Sundaresan, N., & Hartmann, E. (2011). Purchasing competence: A stakeholder-based framework for chief purchasing officers. *Journal of Business Logistics*, 32(2), 122–138.
- Kiggundu, M. N. (1981). Task interdependence and the theory of job design. *Academy of Management Review*, 6, 499–508.
- Kingshott, R. P. J. (2006). The impact of psychological contracts upon trust and commitment within supplier–buyer relationships: A social exchange view. *Industrial Marketing Management*, 35(6), 724–739.
- 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.
- 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.
- 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 relationship marketing. *Journal of Marketing*, 58, 20–38.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557–576.
- Naval Seal Logistics Center Portsmouth. (2014, July). *Guidance for the contractor performance assessment reporting system (CPARS)*. Retrieved October 27, 2015, from [http://dai-global-developments.com/assets/cpars\\_guidance.pdf](http://dai-global-developments.com/assets/cpars_guidance.pdf)
- 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.
- Palmatier, R. W. (2008, July). Interfirm relational drivers of customer value. *Journal of Marketing*, 72(4), 76–89.
- Perkins, W. S. (1993). Measuring customer satisfaction. *Industrial Marketing Management*, 22(3), 247–254.
- Prahinski, C., & Benton, W. C. (2004). Supplier evaluations: Communication strategies to improve supplier performance. *Journal of Operations Management*, 22, 39–62.
- Prahinski, C., & Fan, Y. (2007). Supplier evaluations: The role of communication quality. *Journal of Supply Chain Management*, 43(3), 16–28.
- Purdy, L., & Safayeni, F. (2000). Strategies for supplier evaluation: A framework for potential advantages and limitations. *IEEE Transactions on Engineering Management*, 47(4), 435–443.
- Rizzo, J., House, R., & Lirtzman, S. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly*, 15, 150–163.
- 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.
- 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.
- Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). (2015, April 9). *Implementation directive For Better Buying Power 3.0—Achieving dominant capabilities through technical excellence and innovation* [Memorandum]. Washington, DC: Author.
- 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.
- 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.

## **Acknowledgments**

The author would like to thank the Acquisition Research Program for providing funding support for this research.





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

[www.acquisitionresearch.net](http://www.acquisitionresearch.net)