Collaborative Performance Measurement: Examining and Explaining the Prevalence of Collaboration in State and Local Government Contracts

Anna A. Amirkhanyan
The American University

ABSTRACT

Viewing collaboration as an imperative for public managers, scholars are calling for a better understanding of its origins, prevalence, and impact on organizational performance. The objective of this study is to explore the prevalence and the determinants of collaboration pursued in the course of monitoring government contracts. The theoretical framework proposed in this study explores the effect of several categories of collaboration determinants pertaining to government agencies, contractors, contractual relationships, services, and markets. Data were collected using semi-structured interviews administered to 69 state and local contract managers as well as nonprofit and for-profit contractors in five jurisdictions. Qualitative analysis identifies a variety of collaborative strategies used by agencies seeking vendors’ input and by vendors proposing and negotiating performance monitoring arrangements. Regressions analysing the determinants of collaboration suggest that the latter is more often pursued by nonprofit contractors and vendors with a unique expertise and higher resource dependency. Governments with advanced in-house professional capacity and willingness to collaborate are also more likely to rely on the contractors’ input. Meanwhile, high service measurability, long-term relationships, and dynamic markets reduce the likelihood of collaboration. This study suggests that collaborative performance evaluation presents both challenges and opportunities for effective contract implementation. The responsibilities of monitoring officers appear to extend beyond specifying and enforcing performance standards—they require the skills and motivation to empower contractors and to learn from their input and the professional capacity to evaluate their claims.

INTRODUCTION

Academic inquiries into privatization, performance measurement, and collaboration are analogous in many ways, evolving from somewhat idealized, prescriptive, and normative into more critical, descriptive, and nuanced propositions and approaches (Frederickson and...
Frederickson 2006). Theories of privatization were advanced to rescue traditional public administration from the pitfalls of large and unresponsive Weberian bureaucracies. Performance measurement, in turn, came to address the “limits of privatization”: the principal-agent problems exacerbated by private market failures, inadequate monitoring capacity, and politicized decision making (Lowery 1998; Sclar 1997, 2000; Starr 1987). More recently, awareness of the “dark side” of the performance movement, prone to enforcing generic measurement standards and counterproductive incentive structures (Radin 2006), coincided with the rise of a new “postbureaucratic paradigm,” collaborative public management (Agranoff and McGuire 2003). Although viewing collaboration as an “imperative” (Thomson and Perry 2006, 20) and “an inescapable feature of the future public administration” (Bingham and O’Leary 2006, 165), public management scholars are calling for a better understanding of its origins, prevalence, and impact on organizational performance (Agranoff and McGuire 2003; Imperial 2005; Sandfort 1999).

This research contributes to the growing body of empirical literature on collaboration by studying it in the context of privatization and performance measurement. The first objective of this article is exploratory: to examine the prevalence and the process of collaboration in the course of creating and using performance measurement and monitoring systems in state and local contracts. Specifically, I investigate the prevalence of input seeking, negotiations, and other joint activities initiated by governments and their contractors in order to enhance or modify the process of contract monitoring. The prevalence of collaboration is likely to vary as a function of organizational and contextual factors such as service measurability, the governments’ in-house professional capacity, contractor ownership, and financial autonomy. Hence, the second objective of this research is to articulate and test a theoretical framework explaining the use of these collaborative practices.

The findings of this study suggest that a variety of collaborative activities are employed by agencies and their vendors: monitoring officers seek contractors’ input on performance evaluation, while contractors develop and propose new measures and actively negotiate existing monitoring arrangements. Regressions analysing the determinants of collaboration suggest that it is more often pursued by nonprofit contractors and vendors with a unique expertise and higher resource dependency. Government agencies with higher professional capacity and willingness to collaborate are also more likely to seek and incorporate the contractors’ input. Meanwhile, high service measurability, long-term contracts, and highly dynamic markets reduce the likelihood of joint decision making. The findings of this study suggest that collaborative performance evaluation presents both challenges and opportunities for effective contract implementation. The roles of monitoring officers appear to extend beyond specifying and enforcing a set of performance standards. Instead, these roles require skills and motivation to empower vendors, willingness to learn from their input, and professional and managerial capacity to determine if the vendors’ claims are attuned to contract objectives.

COLLABORATION AND PERFORMANCE MONITORING

Performance Measurement in Government Contracting

Despite the recent decline in the prevalence of new contracts (Hefetz and Warner 2007), contracting out remains one of the most prevalent forms of privatization in the United States
(Brudney et al. 2005). Limited vendor competition and its attendant principal-agent problems (Johnston and Romzek 1999; Lowery 1998) directed the researchers’ attention to the task of contract monitoring (Brown and Potoski 2006; Brown, Potoski, and Van Slyke 2006; Prager 1994). Privatization theorists argue that contract management and accountability “do not take care of themselves” (Johnston and Romzek 1999, 394). In fact, specialized expertise and strong governance structures (Bloomfield 2006) are needed to create an optimal monitoring arrangement that allows comparing “quality and quantity of product or service delivered against contract specifications” (Prager 1994, 179).

Whereas in some studies government agencies are found to be relatively well-equipped for contract management (Brown and Potoski 2003a, 2003b), in other studies, performance information appears to be unreliable, inaccurate, and narrow (Frederickson and Frederickson 2006; Radin 2006). It often focuses on outputs and is subject to diverse interpretations in a world of conflicting interests. When available, standards can also be inadequately enforced, monitored, and interpreted (Heinrich 1999; Prager 1994; Van Slyke 2003). Also, contracted organizations often conceptualize evaluation differently from their principals and engage in activities to show that they are effective, at the expense of evaluation that would allow them to know that they are doing good work (Carman 2007). The latter threatens to result in mission-drift, depprofessionalization, and lower service quality (Van Slyke 2003, 298).

The literature examining the design of optimal contract monitoring and performance measurement arrangement indicates that it is contingent upon multiple factors. When outcome measurability is high, smart-buyer governments tend to rely on “monitoring by proxy” (i.e. by other parties), whereas low measurability and high asset specificity are associated with direct monitoring (Brown and Potoski 2006). In addition, although contracting agencies engage in less monitoring than direct service delivery agencies, governments and their vendors together perform more monitoring than in-house providers do on their own (Brown and Potoski 2006, 327). These findings motivate additional questions on joint monitoring: how and when are joint oversight and “monitoring by proxy” designed and implemented? Who initiates such arrangements? Does joint monitoring result in more effective contract implementation? This study takes the first step to answer these questions. In the next section, I review the literature on collaborative public management which provides a useful framework for analysing joint design and use of performance evaluation procedures.

**Collaboration and Contract Management**

Collaboration has recently gained renewed attention of public management scholars. Although being extensively studied across many disciplines (Bingham and O’Leary 2006), public administration has yet to operationalize this “vague” term (Bloomfield 2006), differentiate it from other related concepts, and, possibly, search for an overarching theory explaining its prevalence, antecedents, and effects. In public management research, collaboration has been broadly defined as a joint multiorganizational arrangement or a set

---

1 It inspired national reforms (Entwistle and Martin 2005) and became a focus of several major scholarly conferences, including the 2006 Collaborative Public Management Conference, Sponsored by the Maxwell School of Syracuse University and The Program on the Analysis and Resolution of Conflicts, and the 2007 American Society of Public Administration spring conference.
of formal and informal interactions that involve (a) jointly institutionalizing new rules, procedures, and structures to govern interorganizational relationships; (b) making joint decisions about fiscal and personnel management; (c) solving problems that cannot be solved by single organizations within the existing informational, financial, time, and market constraints; and (d) creating new public value or making a joint discovery (Agranoff and McGuire 2003, 20; Bazzoli et al. 1997, 540; Imperial 2005, 297; McGuire 2006, 37; Selden, Sowa, and Sandfort 2006; Thomson 2001; Thomson and Perry 2006).²

The concept of collaboration embraces both higher level, formal institutional arrangements, such as networks or contracting dyads, and specific activities pursued by public managers (Bingham and O’Leary 2006). It involves self-interested bargaining and relationships based on commitment to larger interests (Thomson and Perry 2006). Collaboration presumes a deviation from hierarchical decision making; it requires participation of several actors in the governance process and some degree of trust in the partners’ competencies that are “unlocked” through collaboration (Entwistle and Martin 2005, 2004).

Theorists propose that collaboration is determined by a variety of factors such as levels of interdependence, past efforts to collaborate, and resource dependence (Thomson and Perry 2006). Research in the field of health and human services shows that services are provided collaboratively by at least 50% of partnerships (Bazzoli et al. 1997). Collaborative efforts are common in emergency operations and other fields (Good 2003; Kapucu 2006). Limited studies on the outcomes of collaboration show that collaborative activities may have a positive effect on early care and education (Selden, Sowa, and Sandfort 2006), whereas their absence may create inefficiencies in service delivery (Sandfort 1999).

The notion of collaborative performance evaluation may be rooted in implementation studies that stress the role of mutual adaptation, learning, and feedback as key components of policy evaluation, since the latter is “a conscious attempt to generate and learn from policy-relevant feedback” (Pressman and Wildavski 1984, 227). More recently, the concept of collaborative contract management was advanced by DeHoog (1990) who suggested moving beyond the “stated preferences” in contract implementation and acknowledging that contract effectiveness is often explained by individual relationships. While the competition model of contracting promotes efficiency and bureaucratic procedures, the negotiation and cooperation models are based on the idea that the government and the contractor are equal partners who “try to anticipate the implementation problems, yet knowing that they cannot write all possible contingencies into the contract” (DeHoog 1990, 330). In cooperative contracts, monitoring officials work jointly on overcoming obstacles, using common language, and maintaining relationships rather than enforcing rigid standards and looking for other providers. As a result, the contractors’ opportunistic behavior and the transaction costs are minimized, and performance is influenced by professional and informal rather than bureaucratic accountability pressures. Extending the role of collaboration to the vendors’ performance evaluation, Page (2004, 599) proposes that collaboration may be central to the ability to collect reliable data and measure results in the field of human services:

“A human services collaborative’s capacity to measure results depends on agreement among the collaborators and their authorizers about what to measure, as well as its ability to track

² According to some authors, collaboration is synonymous with cooperation or partnership (Gazley 2008; Kapucu 2006), whereas others differentiate these concepts (Agranoff and McGuire 2003).
changes in the results and to use the findings to identify priorities for improvement in the future. Joint agreement on measurable results can foster common ground among collaborators, enhance the credibility of their initiatives, and enable them to refine their aims and strategies as performance data become available.”

Some empirical evidence on collaboration in contract management is provided by Bovaird (2006), Imperial (2005), and Johnston and Romzek (1999, 2002). Johnston and Romzek (1999, 387, 391) found that accountability relationships in state-level contracts were based on mutuality: although the state dictated the terms of the contract and monitored compliance, there was a recognition of the contractors’ flexibility, discretion, negotiation, and collaborative problem solving as the program moved forward. Johnston and Romzek (2002) note that “[i]f good will and trust exist between the agency and the contractor, then the parties often proceed with the understanding that performance expectations will be negotiated, and if necessary adjusted, as the program unfolds” (Johnston and Romzek 2002, 436). Collaboration has also been found to enhance the existing monitoring and enforcement efforts by creating systems in which partners agree on the sampling design and monitoring protocols (Imperial 2005). Bovaird (2006) analyses a case of “imaginative partnership” in which collaborative efforts result in innovative record keeping, information technology, and other performance monitoring practices. Thus, there is some empirical evidence of collaborative performance monitoring in government contracting. In the following sections, I examine the process, the prevalence, and the determinants of such activities in a sample of state and local contracts.

THEORETICAL FRAMEWORK

The Prevalence and the Process of Collaboration

In this study, collaboration involves joint, coordinated, or connected activities of monitoring officers and their vendors conducted in order to sustain, modify, and enhance contract measurement, monitoring, and evaluation. The first research question explores whether collaboration is present in contract monitoring and how it happens. Figure 1 provides
a simple way to conceptualize the prevalence of collaborative activities. First, despite the importance of performance evaluation, some government contracts may not be monitored. The latter, by definition, would preclude the parties from doing so “collaboratively.” Contracts involving a certain degree of performance monitoring may be represented on a continuum from highly collaborative models to low or nonexistent levels of collaboration in what are essentially “compliance-based” or “top-down” relationships (Agranoff and McGuire 2003). This variation can also be conceptualized as a dichotomy between the use of unilaterally designed and implemented monitoring systems (Bazzoli et al. 1997) and those involving some degree of collaboration—flexibility and dialogue between the agency and the vendor.

Specific collaborative activities in contract monitoring are extremely diverse and, notably, different from the traditional POSDCORB. They include developing ways to manage complexity, information seeking, adjustment seeking, interpretation of standards and rules, changing policies, learning about constraints, reading each others’ signals, creating shared meaning, consulting, empowering, and bargaining (Agranoff and McGuire 2001, 2003; McGuire 2006; Page 2004). Information is the key ingredient of these activities (Agranoff and McGuire 2003). Collaborative performance measurement may be especially concerned with the type, sources, diversity, and quality of information collected to monitor service delivery. In the context of this study, it is interesting to explore these activities while paying specific attention to the party that initiates or actively promotes collaboration. This, certainly, includes the possibility of both parties initiating and responding to collaborative actions. Collaborative activities may be initiated by government agencies that actively seek vendors’ input on performance measurement. This can happen in cases when a new contract is developed or in the course of contract implementation when a monitoring officer lacks knowledge within a specialized field or has concerns about the vendor’s performance. Vendors, in turn, can also initiate negotiations and discussions pertaining to performance expectations, as well as the sources, the method, the type, the format, and the quality of information collected to assess their work. This may occur when performance standards are perceived by the vendors as inadequate, disruptive, difficult to comply with, or misrepresenting their actual performance. Vendor’s feedback may be incorporated fully or partially by modifying the existing performance monitoring tools. Collaborative activities may also involve joint research into professional norms and standards in the field, as well as clarifications of service-delivery mechanisms.

**Determinants of Collaboration**

The second objective of this study is two-fold. It involves putting together and testing a theoretical framework explaining the prevalence of collaborative activities in the course of contract monitoring. The determinants proposed here are grouped into five categories.

**Service Characteristics**

The role of transaction costs in contract design is well-established in the privatization literature (Brown and Potoski 2004; Williamson 1981, 1996). Compared to the top-down enforcement, elements of collaboration—clarifications, input seeking, feedback, and negotiations—require additional time and resources aside from the usual administrative costs (Hefetz and Warner 2007; Thomson and Perry 2006). The decision to incur these costs may be rooted in service measurability. Service measurability can be conceptualized
using the distinction between “hard” services, with tangible and easily quantifiable end results, and “soft” services that are harder to measure due to multiple, complex, and long-term outcomes. Wilson (1991) offers a more nuanced classification of services into “production” (outputs and outcomes easily observed), “craft” (unobserved outputs), “procedure” (unobserved outcomes), and “coping” (outputs and outcomes unobserved). Facing greater informational asymmetry and risk of failure in the case of soft services in procedural or coping agencies, managers may be forced to accept higher levels of management complexity (Johnston and Romzek 1999). They may incur transaction costs and engage in collaborative discussions in order to develop mutually acceptable performance expectations. Eventually, the need for collaboration and the associated transaction costs will decline as performance standards are refined, and trust is developed (DeHoog 1990). Hence, with the relationship length kept constant, services with lower measurability may be associated with higher collaborative activity.

**Contractor and Market Characteristics**

Collaboration is a two-way street, and both vendors’ and purchasers’ characteristics affect the likelihood of pursuing collaborative activities. Contractor ownership, in particular, has been extensively studied in the context of organizational performance (Amirkhanyan, Kim, and Lambright 2008; Heinrich and Fournier 2004; Morley 2006). The theories of nonprofit enterprises argue that due to the legal nondistribution constraint nonprofits lack incentives to raise prices; they are responsive, trusted by their clients, and more likely to serve as a safety net for vulnerable populations (Hansmann 1980, 1996; Prager 1994). Relying on government funds, nonprofits are to a lesser degree influenced by the economic markets, and their governance structures reinforce internal oversight (O’Regan and Oster 2002). In contrast, many point to the clash of public and for-profit service-delivery philosophies: one focuses on individual outcomes and the other on the financial bottom line (Dias and Maynard-Moody 2006). For-profit status is also perceived to present additional risk of abuse warranting more aggressive control by public agencies (Morley 2006). Hence, governments may use nonprofit ownership as a signal of honesty and rely more readily on their input. Meanwhile, the lack of trust, more prevalent in the government-for-profit dyads, may be associated with a top-down model of enforcing performance standards.

The contractor’s resource dependency may also affect the prevalence of collaboration. If a vendor’s financial health depends largely on a single contract, ensuring the purchaser’s satisfaction becomes particularly important. This may determine the contractors’ desire to seek adjustments in performance evaluation. Contractors may volunteer additional data to enhance the government’s perception of outcomes, they may provide interpretations of performance data, and negotiate reporting requirements in order to reduce the administrative burden and focus on service provision. In summary, financial dependency may encourage the vendors to provide feedback, and that, in turn, will increase the chance of incorporating this feedback into the evaluation process.

Many private organizations conduct internal performance evaluations allowing them to react promptly to emerging problems. The existence of internal measures reflects some degree of management capacity and commitment to performance improvement which may facilitate better accountability. Contractors using internal performance evaluation may seek coherence between the two systems and try to reduce redundancy by engaging in negotiations with the government agency.
In figure 2, two factors account for stability: market competition and the overall vitality of the organizational environment. In a competitive market, purchasers and their vendors have less incentive to incur transaction costs by undergoing mutual adaptation. Competition encourages efficiency and warrants bureaucratic procedures in contract management (DeHoog 1990). Relying on the competitive pressures to maintain service quality, purchasers may forgo sophisticated performance evaluation and may have less need to collaborate in the process. A similar rationale can be applied to the overall dynamism of organizational environment. Whereas some fields are quite stable, others undergo rapid changes in service needs, technology, or funding. Dynamic environments may discourage partners from investing time and effort in the relationships, whereas stability may motivate more mutual adaptation and collaboration. As Milward and Provan (2000) note, “[s]tability promotes a belief that cooperation will have a reasonable payoff” (2000, 376).

**Government-Vendor Relationship**
A collaborative relationship between a public agency and a contractor may be developed after a lengthy period of observing each other’s work ethic and developing interpersonal ties. Long-term relationships presume that the two sides have forgone alternative options;
they reflect complacency with mutual performance which may affect the likelihood of collaboration. Trust developed in the course of the relationship and changes in the environment may require some adjustment in performance expectations and stimulate the use of consultations, negotiations, and other forms of participatory performance evaluation. Hence, long-term relationships and the associated perceived goal congruence and trust may result in higher prevalence of collaborative activities.

Some contract monitoring arrangements may necessitate collaboration whereas others may discourage it. In the course of monitoring, governments may (a) rely on self-reported data presented by vendors, (b) directly collect data through interviews or inspections, and (c) rely on third-party evaluators, such as professional associations and accreditation agencies. When performance measurement is delegated to the contractor, vendors are de facto engaged in the evaluation process. They will be likely to discover problems and choose to clarify and negotiate various aspects of this process. Meanwhile, when performance is observed or inspected directly, monitoring officers will be more likely to use unilaterally developed procedures.

**Government Characteristics**

Government agencies’ in-house professional capacity to deliver services may influence contract design (Prager 1994) and the likelihood of collaboration. Milward and Provan (2000) argue that principals involved in production are more effective than those only engaged in the governance process. According to DeHoog (1990), common professional standards increase the likelihood of a productive dialogue since shared values encourage communication. The government’s internal capacity may be more prevalent in contracts involving joint service provision. It may also be associated with reliance on professional accountability mechanisms based on widely recognized and mutually acceptable standards and norms (Johnston and Romzek 1999; Page 2004). Both factors can encourage a collaborative dialogue on performance expectations and monitoring.

**Respondent Controls**

Collaboration is more likely to occur when participants are willing to collaborate and if the purchasing officers or the vendors perceive the contractor’s input as useful and important in the process of performance monitoring. In some cases, governments may approach the contract as a purely economic transaction in which the “buyer” has full control over monitoring. In such cases, contract managers may be less likely to seek and incorporate vendors’ input, whereas vendors who perceive their role as implementation will refrain from negotiation.

In addition, since manager’s capabilities are central to effective contracting (Brown, Potoski, and Van Slyke 2006), this study controls for the respondents’ work and contract management experience. On one hand, past contracting experience may motivate public managers to enhance evaluation using participatory mechanisms. On the other hand, the field may be the ultimate determinant of collaborative activities: the smart-buyer governments may evaluate the costs and the benefits of collaboration and apply it selectively based on the contextual factors, rather than their capabilities. Findings of Gazley (2008) support the first explanation: experience and familiarity with other sectors, among other factors, predict the managers’ receptivity to collaboration.
METHODS

Data

This study uses a combination of qualitative and quantitative methods. Data were collected using a semistructured interview instrument, administered to contract monitoring officers in state and local governments, and the managers of nonprofit and for-profit contractees. The instrument included questions with nominal or ordinal response categories, and those requiring descriptive explanation and, hence, producing data that were appropriate for both qualitative and quantitative analyses. Documentation pertaining to the contract (e.g. requests for proposals, notices of awards, and amendments) was also acquired. Data collection was conducted in the District of Columbia, three adjacent counties, and one adjacent state. In all but one jurisdiction, the information on current service contracts and the contacts of monitoring officers and vendors were obtained using search engines available on the procurement office Web sites. A purposive sample of monitoring officers was selected within each jurisdiction with the purpose of maximizing the representation of service fields, service measurability, award amounts, vendors’ ownership, and other factors. When a single contract officer was assigned to monitor several contracts, respondents were asked to choose and discuss the most typical contract they oversaw. A similar strategy was used to select the vendors. A purposive sample of contractors was drawn attempting to (1) avoid two-sided representation of the same contract, (2) maximize the diversity of service areas, and (3) ensure that the vendors and the contractors were proportionally represented from each location. Most vendors in the sample operated either locally or across several neighboring jurisdictions. Two pretests were conducted as a result of which the questionnaire was modified. The final sample included 69 responses: 39 in government agencies and 30 in private organizations. Table 1 shows the distribution of service areas in the sample. The proportion of for-profit vendors in the sample is higher than that of nonprofit vendors. Sixty-seven percent of monitoring officers discussed a contract with a for-profit organization, whereas 33% discussed contracts with a nonprofit. In the contractor sample, 63% were for-profit and 37% were nonprofit.

3 Please contact the author for a complete copy of the interview protocol.
4 These jurisdictions provided convenient access for the author and allowed conducting most interviews in person. Although the geographic area is limited, the unique location of the DC metropolitan area provides an opportunity to access multiple jurisdictions in Maryland and Virginia. Population in the jurisdictions included in the study ranged from 195,000 to 5,600,000. The proportion of Whites ranged from 38% to 80%, and the median household income ranged between 43,000 and 77,000.
5 In one jurisdiction, a telephone book was used to reach the subjects.
6 The total of 92% of government respondents were contacted based on internet listings, whereas 8% were found using snow-ball sampling or referrals. The “technical” or “program” officers were interviewed in 96% of all public-sector interviews. In the remaining 4% of cases, the interviews were conducted with the procurement officers who were knowledgeable about monitoring practices. Each subject participated in one interview, most of which were completed within 1 h.
7 This study did not pursue the objective of interviewing government officers and vendors associated with the same contract. Rather, two purposive groups of respondents from government and vendor populations were selected separately. While such approach prevents us from examining two sides of joint contractual arrangements, independence of observations allows including them in a single regression model without biasing the result.
8 The exceptions were national companies that also provided services locally.
Analysis

The analysis presented in the next section begins with an overview of the monitoring procedures used in the sample. Using questions one, two, and three in Appendix 1, I analyse the monitoring practices and examine the prevalence of measures reflecting different aspects of contractor performance. The analysis then proceeds to the examination of collaborative practices using questions four through seven in Appendix 1. These questions were used to create four dummy variables (“asking for input,” “contractor negotiation,” “input incorporated,” and “communication affects performance”), each coded as one for affirmative and zero for all other responses. I report the prevalence of these four activities and analyse descriptive responses to these questions. Finally, quantitative analysis is used to examine the impact of various antecedents on the four measures of collaboration listed above. The log of odds of four dependent variables was regressed on the independent variables, described in Appendix 2, in order to test the effect of various organizational and environmental factors on the likelihood of pursuing these activities.9

Limitations

Numerous limitations are associated with the chosen research design. Among them are threats to generalizability due to the limited number of jurisdictions. The findings of this study may be applied more readily to locations with higher median income and those close to large metropolitan areas. Although the reliability of data collection was improved by using a common data collection protocol, future replications of this study in other locations may help enhance its reliability and external validity.

9 A dummy variable was added to indicate if the respondent was a monitoring officer or a vendor.

Table 1
Service Areas Included in the Sample

<table>
<thead>
<tr>
<th>Service Area</th>
<th>Government</th>
<th>Contractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information technology</td>
<td>7.7</td>
<td>6.7</td>
</tr>
<tr>
<td>Long-term care</td>
<td>17.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Medical, nursing care, health management</td>
<td>12.8</td>
<td>6.7</td>
</tr>
<tr>
<td>Mental health, psychological consultation, arts therapy</td>
<td>10.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Construction, maintenance, public works</td>
<td>15.4</td>
<td>10.0</td>
</tr>
<tr>
<td>Programs for women and children</td>
<td>2.6</td>
<td>10.0</td>
</tr>
<tr>
<td>Consulting, evaluation, and management training</td>
<td>23.1</td>
<td>20.0</td>
</tr>
<tr>
<td>Criminal justice</td>
<td>2.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Environmental (planting, plant control)</td>
<td>5.1</td>
<td>6.7</td>
</tr>
<tr>
<td>Food supply (and quality) monitoring</td>
<td>2.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Animal care</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Substance abuse, homelessness</td>
<td>0.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Janitorial</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Translation</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Recreational (camps, dance lessons)</td>
<td>0.0</td>
<td>6.7</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Due to methodological issues, past contractor performance, which may determine collaboration, is unaccounted for in this analysis. Respondents’ agreement to participate in this study was often predicated upon abstaining from direct discussions of performance. In addition, soliciting responses on perceived performance from both monitoring officers and vendors could result in systematic differences necessitating stratification of the analysis. The latter would reduce the sample size. Two measures presently used in the analysis may serve as imperfect proxies for contractor performance: relationship length and perceived contractor expertise. Although being narrow, both measures reflect complacency with the relationship and acknowledgement of the vendors’ capacities.

Finally, a separate concern for validity in the qualitative analysis is caused by overrepresentation of the for-profit vendors. Whereas the effect of vendor ownership on monitoring practices has not been extensively examined in the empirical literature, for-profit ownership may be associated with higher service measurability, which, as determined in the analysis below, is less likely to be monitored collaboratively. Thus, it is possible that a better representation of nonprofit providers could have revealed a wider variety of collaborative practices.

THE PREVALENCE OF COLLABORATIVE PERFORMANCE MEASUREMENT

Overview of Contract Monitoring Procedures

Since this article focuses on collaborative activities undertaken in the course of performance monitoring, this section provides a context for the main findings by describing the monitoring procedures and measures utilized in the sample. Figure 3 shows the proportion of monitoring officers and vendors who indicate that the contract involved (a) self-reporting by the vendor, (b) direct government monitoring, and (c) third-party monitoring. Self-reported measures appear to have the highest frequency: 77% of contractors and 92% of monitoring officers report using them. Government monitoring was reported in 57% and 85% of contracts by private and public subjects, respectively. Third-party monitoring was used only in 30% and 54% of contracts discussed by vendors and contract officers. The prevalence of each approach is higher among government respondents.

Public services have multiple outcomes, and a variety of performance measures can be used to capture different aspects of performance (Behn 2003; Berman and Wang 2000; Boyne et al. 2005; Edwards and Thomas 2005). In this study, respondents were asked to report which aspects of the vendors’ performance were monitored (question 2 in Appendix 1). Table 2 shows the proportion of contracts that use each performance measure. Generally, similar patterns were reported by the contractors and government officers: majority of respondents reported measuring service quality, timeliness, continuity, vendors’ workload, and also relied on informal monitoring. Fewer contracts evaluated the impact of services on clients and client satisfaction or used quantitative and qualitative performance indicators. The lowest proportion of respondents in each group reported eval-

---

10 The quantitative analysis controls for ownership.
11 In this study, contract monitoring and performance measurement involve the design, collection, and evaluation of data describing various aspects of the vendors’ activities.
Evaluating costs or cost-effectiveness, reputation, ability to provide services without discrimination, to specify detailed procedures for service delivery, or to tailor performance measurement to the contracted organization. Once again, public managers are more likely to report monitoring performance compared to the vendors. This may have several explanations. First, contractors may not be fully aware of the extent of government monitoring, particularly in cases when governments perform independent inspections. The second explanation may be related to the vendors’ perception of the quality of government monitoring. Some vendors can be skeptical of government’s ability to monitor performance. When asked if the government agency monitored service quality, some responded “They think they do! . . . But I really don’t think they do it . . .” or “I don’t know. They do something to that end, but I don’t think they can actually evaluate quality.” Possibly, whereas contract officers reported on their propensity to monitor performance, contractors may have been commenting on the governments’ propensity to monitor effectively. Finally, social threats

---

12 Low prevalence of cost monitoring was determined by a large proportion of fixed-cost contracts in the field of human services. In such cases, contracting officers reported receiving and checking the billing documentation, rather than evaluating the cost-effectiveness of service delivery.
to validity and overreporting of monitoring due to evaluation apprehension cannot be eliminated.

In addition to these quantitative data, respondents were asked to describe the procedures used to measure, monitor, or evaluate contractor performance. Regular progress reports are among the most common performance evaluation techniques. These include information on staff credentials, service delivery logs, changes in client status, narratives, and comparisons of vendor performance indicators to the standards set in the contract. Many subjects also communicate directly, formally, and informally, to review the proportion of completed tasks and to discuss the ongoing problems. For example, a monitoring officer of a human-service nonprofit illustrated the importance of informal monitoring: during a conversation with a vendor, she inadvertently found out that a client, presumed to be receiving the service, had in fact been incarcerated, while the vendor continued billing the agency.

Interviewees also report relying on direct government monitoring. Site visits, often unannounced, are conducted at the time of service delivery in order to observe the workflow, examine documentation, staff attendance, client satisfaction, and vendor governance structures. While some monitors actively seek customer feedback through surveys or focus groups, others presume that the involved parties would come to them should a problem arise. Contracts with nonprofit vendors are associated with the use of ombudsmen and boards of trustees in the monitoring process. Additional activities include monitoring billing information, making sure that the contractors comply with the laws and regulations in the field, and reviewing information provided by the third-party monitoring agencies.

Although involving ample evidence of collaboration, presented in the following sections, these monitoring activities incorporate elements of conflict and frustration previously documented in the contracting research (Van Slyke 2003). Discussing their inability to effectively measure the vendors’ performance, respondents mention their lack of training, capacity, internal infrastructure, and resources:

<table>
<thead>
<tr>
<th>Performance Measure/Indicator</th>
<th>Government (% “yes”)</th>
<th>Contractors (% “yes”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs/cost-effectiveness</td>
<td>64.1</td>
<td>46.7</td>
</tr>
<tr>
<td>Quality</td>
<td>94.9</td>
<td>66.7</td>
</tr>
<tr>
<td>Workload</td>
<td>87.2</td>
<td>63.3</td>
</tr>
<tr>
<td>Impact on clients</td>
<td>79.5</td>
<td>56.7</td>
</tr>
<tr>
<td>Client satisfaction</td>
<td>84.6</td>
<td>53.3</td>
</tr>
<tr>
<td>Equitable delivery of services</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>Compliance with laws/regulations</td>
<td>64.1</td>
<td>43.3</td>
</tr>
<tr>
<td>Timeliness</td>
<td>94.9</td>
<td>66.7</td>
</tr>
<tr>
<td>Disruptions</td>
<td>97.4</td>
<td>73.3</td>
</tr>
<tr>
<td>Process specified in details</td>
<td>69.2</td>
<td>43.3</td>
</tr>
<tr>
<td>Quantitative indicators</td>
<td>76.9</td>
<td>60</td>
</tr>
<tr>
<td>Qualitative indicators</td>
<td>82.1</td>
<td>66.7</td>
</tr>
<tr>
<td>Informal monitoring</td>
<td>94.9</td>
<td>70</td>
</tr>
<tr>
<td>Measures tailored to organizations</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>Reputation</td>
<td>64.1</td>
<td>56.7</td>
</tr>
</tbody>
</table>

*Note: Performance Measurement, PM.*
I haven’t really been given the resources, staff, the know-how, any of that, to be able to really say I am doing service to the taxpayers. Do I take the check sheet that the office of procurement gives [me] and says “Check this sheet!”—do I do that? Yes. But do I really have the time, with everything else, to be able to make sure that the [contractor’s employees] show up on time? No, I am not on site . . .

Such concerns are exacerbated by the organizational culture expressed by monitors and vendors:

We don’t come from a culture of sharing and collaboration. It’s a culture of rivalries and jurisdictions. It’s not even the matter of silos, it’s really affirmative posturing people take around rivalries, budgeting and jurisdictions. The [staff] has some responsibilities, us, monitors, have, sort of, responsibilities around that, and so does the [procurement officer]. So there is a lot of finger pointing.

That’s the nature of the beast. The contractors are forprofit and the government agencies are all for procedures. These two schools clash. Sometimes they don’t have the authority to make modifications, sometimes the personnel are not qualified to do their work.

Some contractors report having excessive reporting requirements, whereas other vendors claim that they are not being monitored or are monitored very loosely:

Perhaps they are monitoring the quality informally, but we don’t think they are really doing that . . . as a procedure.

No, as hard as it is to believe, they do nothing.

They do. But to no great extent.

They know what service we provide, and they could respond to complaints, but other than that they don’t monitor it.

In sum, monitoring activities used in the sample varied greatly in their form and complexity.

The Prevalence and the Process of Collaborative Activities

Table 3 shows the prevalence of affirmative responses to four questions reflecting the contractors’ involvement in performance measurement. In most cases, the side initiating a specific collaborative activity is more likely to report it: vendors are more likely than governments to report negotiating and discussing how their performance should be measured and evaluated (57% versus 41%), whereas government agencies are more likely to report seeking vendors’ input (28% versus 17%). Overall, seeking contractors’ input is less prevalent than the contractors’ attempts to negotiate performance evaluation. Thus, contractors appear to be more active in seeking joint decision making. Notably, monitoring officers report incorporating the contractors’ feedback in as many cases as they report seeking feedback. Meanwhile, according to the vendors, government agencies incorporate their feedback about as frequently as the vendors engage in negotiations. Although government agencies are quite reluctant to admit that they incorporate vendors’

---

13 Title removed for confidentiality.
14 Professional titles removed for confidentiality.
15 Job title removed for confidentiality.
input (28% versus 50%, for governments and vendors, respectively), they are more likely to report that their overall communication with the vendors produces changes: over half of respondents in both groups (53% and 61%) report that overall communication with the opposite party influences the measures that are eventually used. In sum, although the components of collaboration are reported differently by each side, about half of all respondents claim that monitoring practices are modified as a result of communication. This is consistent with the findings in the field of health and human services (Bazzoli et al. 1997).

Analysis of qualitative responses helps identify a variety of collaborative performance measurement strategies used by both parties. First, I describe some government-initiated strategies of joint examination of performance standards. Second, I discuss the vendors’ approaches to developing and proposing performance measures and negotiating existing monitoring systems. Finally, I discuss the instances of unilaterally designed and implemented monitoring systems, that is, the cases of “collaborative abstinence.”

### Government-Initiated Joint Examination of Performance Standards

A joint examination of performance standards and measures conducted by government agencies and their vendors is one of the most common forms of collaboration initiated by monitoring officers. This process commonly takes place at the time a contract is awarded, but it can also be a part of the formal periodic performance review process:

> We bring in the [vendor] and we talk about performance measures: stuff that we should look at and stuff that doesn’t really have a true impact.

> As a joint effort, we tried to assess the value of the data we have collected.

> We validate with the contractor that this is something they would be able to do.

> Working with the contractor helped us to determine our performance measures.

> They asked us for certain information and if we could give it to them. It certainly was not an authoritative approach, just saying: “this is the information that we need.” We sat down and talked about what kinds of information we should be looking at. It was a joint effort.

Discussions with vendors may cover specific performance standards such as the rate of timely responses, service use, and ways of demonstrating “the informal piece” of performance. The discussion of performance standards often accompanies a broader discussion of service delivery, particularly in cases when new programs are developed. Vendors are

<table>
<thead>
<tr>
<th>Collaborative Activities</th>
<th>Government</th>
<th>Contractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractor negotiates and discusses PM</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Government asking contractor for input</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Contractor’s input is incorporated</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Communication affects PM</td>
<td>24</td>
<td>16</td>
</tr>
</tbody>
</table>

Note: Performance Measurement, PM.
asked to help “shape the service,” “create a policy,” and come up with the associated reporting requirements.

Informal communication with the vendors also commonly results in interpreting performance information or discussing the status of a specific case. One respondent described this process: “Sometimes they give me some information, and it’s sufficient. In other cases, I ask for some additional data and we discuss what we can do to get it. I listen to them. It’s not a standardized process. It’s informal.” Monitoring officers also investigate external complaints or unsatisfactory trends. In such cases, vendors “try to explain why something wasn’t done and ask why we feel it’s inadequate. Sometimes the explanation is a viable one, and we accept it.” In one case, a system developed to track the task completion had to be adjusted for clients who had no telephone to avoid misrepresenting the vendor’s productivity. In another case, a vendor working on equipment maintenance routinely failed to “close the work orders” in a shared electronic system. While the vendor requested full compensation for such cases, the monitoring agency followed the terms of the contract and applied full compensation to the cases when completion was formally recorded by the work order system. In a different contract, an agency investigated the discrepancies with the data supplied by the previous contractor. Eventually, the government and the new contractor went through a “reconciliation process to make sure the numbers matched the ones that were previously generated.”

Vendors Developing Performance Measures and Negotiating

A number of vendors report developing and sharing new performance monitoring procedures with public agencies. Often, they are intended to replace the existing measures that are not perceived as “workable” or “feasible.” Contractors also choose to provide more information than originally requested by the agency. This information may refer to the contractors’ overall organizational performance (such as staff turnover rates), as well as data intended to provide a broader picture of client outcomes. Contractors also supply performance standards and the best practices from their fields and based on their research. Reflecting on vendors’ feedback, monitoring officers argue that the vendors’ contributions help them to expand their perspectives:

In the process, they felt the need to present additional information that helped us see the bigger picture.

They were willing to make suggestions, which have been accepted, and changes were made.

After the first couple years, it was really about enrollment. We were not looking at the population itself. But we said “Let’s take a different approach: see if we can change behavior.” We went back to research, did the assessment and said to our clients [government agency—A.A.] “Let’s look at the major health issue that is affecting this city and use it to increase awareness.

Informally they’ve always listened to our narrative stories. But they didn’t change the formal system.

In several cases, vendors report volunteering performance information in the absence of any monitoring tools used by the monitoring agency. For example, training organizers hold debriefing sessions and share results with the government agency, while mental health professionals conduct assessments for internal purposes and disclose findings to their monitors.
Some vendors suggest submitting progress reports that have not been originally required: “We said: “There has to be some reporting” . . . I said: “Come to our board meetings. We want you to know where we stand.” Certain vendors pursue more aggressive strategies to make sure their perspectives on performance evaluation are taken into consideration:

We were not asked. But we took it upon ourselves to tell them.

We have forced them to make this contract fit our program over the years . . . . We have showed them that it is not equitable by the nature of [clients] to use attendance and, if they do not attend, kick them out or penalize us by taking away the money. So, in terms of performance measures, you need to adjust the contract to the population. This is the reality of the population, of these [clients]. They’ve tried to cookie cutter before . . . . So, they have tailored it to us, but not willingly.

Monitoring officers also discuss the instances of being pressured into adopting certain practices:

In [our field] you don’t get a lot of continuity. There is no investment when we go in and say you have to do your healthy checks stuff: the ear, the nose, etc. When eight hours later you have a brand new workforce, you can’t go and say it again. Eventually, you say “Whatever, document it however you want, I couldn’t care less.” Literally, my workforce changes in 8 hours.

Contractors report the need to negotiate performance standards in cases when government expectations appear to be unreasonable or hard to comply with. Vendors negotiate their workload by demonstrating limited capacity and questioning their responsibility to collect certain data. Monitoring officers, in turn, acknowledge the lack of infrastructure and, in some cases, modify requirements to fit the vendors’ capabilities.

They will call and say we don’t think this is all that reasonable. And I say—well, tell me why . . . . Sometimes, I say—you’ve come up with a really compelling case. And I say, OK, provide me this and this and I will push through the fact that this satisfied the requirement.

I am really sympathetic with the providers . . . in terms of criminal record check. Because God only knows how they do their due diligence. My provider is calling me and saying: “Yeah [sic], of course, you are gonna [sic] describe the due diligence in your contract like super-duper, but you are not doing it. And you are expecting us to do that?” The mom-and-pop shops are really having trouble dealing with this financially. Because they don’t have the infrastructure and the “know-how.”

We have initiated most of the communication if we felt performance measurement wasn’t serving our mission well. But they just made a change that is completely unattainable for us since the beginning—we told them that we will never be able to reach the goals they set.

The need for negotiation is often determined by the service-delivery conditions unforeseen in the contract. For example, a vendor working on a construction project was unable to follow the industry guidelines due to weather and soil conditions that threatened to obstruct their expensive equipment. The standard was lowered when the monitoring officer examined the site and concluded that it could be done with no impact on quality. In a different contract, the vendor argued that by asking her to operate in random geographic
locations, the government inadvertently undermined her efficiency. The monitoring officer did not change the referral process but took this into consideration in the evaluation process. Another contract involved a successful negotiation involving an employee who had not passed the background check but was perceived to be an excellent worker. Several vendors expressed the need to negotiate in cases of the clients’ unwillingness to receive the service or the subcontractors’ failure.

Poor performance measures are among the most common causes for negotiation. Some vendors engage in negotiations when performance evaluation fails to capture the full complexity of their performance. A human service organization was asked to use a software developed by the US Department of Housing and Urban Development which, according to the vendor, incorporated only narrow, quantitative measures of client well-being. To complement these measures, the vendor added some qualitative feedback in his communication with the agency. Referring to the compulsory reliance on federal guidelines, he noted sarcastically that the involved parties had to be bilingual and “*speak English and HUD*.” Another vendor described the case of a government agency’s participation in a “*Managing for Results*” initiative. An external monitor insisted on using the employee head count as a performance measure:

One measure they pulled was head count for the staff, and we didn’t feel that was appropriate. Head count does not really reflect whether or not the services are being delivered. We’ve had discussions on this and they said, really, *[there is]* nothing we can do about it. The state can’t change it—they would need to amend the contract. But I hope to see some of these things get cleared up in the next RFP cycle. They understand that this is really not a good measure, they wouldn’t have selected it—it was selected by somebody else . . .

Other contractors question the validity of measures that are not tailored to their context. One contract delivered services in a rural area while using the standards that were more appropriate for metropolitan areas. Another contract required excessive medical paperwork for children participating in short-term recreational activities. Contractors also report ample examples of clarifications and discussions pertaining to the government’s expectations, clarification of terminology, as well as the government’s use of performance information.

“Collaborative Abstinence”

A number of agencies employed top-down approaches in the course of performance monitoring claiming that there was no give and take in ensuring accountability: “*We determine the criteria and the process. [Contractors] don’t have any say in that.*” Some monitoring officers felt bound by the terms of the contract or professional guidelines. Another respondent believed that the initial research during the request for proposals (RFP) phase was sufficient for ensuring effective monitoring. Several respondents argued that high service measurability and existence of “*cut and dry*” service outcomes eliminated the need for joint decision making. Notably, such responses were obtained in a variety of fields, including long-term and nursing care. Some vendors confirmed that their participation was discouraged. When asked if they attempted to negotiate performance evaluation, one respondent noted: “*No, I don’t. [The agency] would be opposed to it.*” Finally, vendors reported collaborative abstinence in cases when little or no monitoring was done, or no feedback on performance was provided: “*They never contacted me to discuss the results of the survey, never came back to say, ‘We acknowledge your good work’.*”
Summary

There is considerable evidence of joint decision making related to performance measurement and monitoring in government contracts. Agencies rely on vendors’ input, and vendors in turn do not shy away from providing it. Vendors share their perspectives by negotiating the procedures that are perceived as inefficient and adding new tools that provide a more complete picture of their performance. Monitoring officers not only listen to the providers’ concerns but also take time to consider the effect of suggested changes and choose to follow some suggestions and forgo others. What explains the governments’ adoption of contractor feedback and the prevalence of other collaborative activities in state and local contracts? In the next section, quantitative analysis is conducted to answer this question.

DETERMINANTS OF COLLABORATIVE ACTIVITIES

Table 4 shows results for four logistic regression models examining the effect of various factors on the likelihood of collaborative activities. First, I examine the goodness of the fit statistics. Models 3 and 4 show a good fit: chi = 36.6703 and p value = .0058 in model 3 and chi = 31.628 and p value = .0243 in model 4. Whereas the Wald test shows significant results for several predictors in models 1 and 2, their explanatory power is poor: chi = 23.2808 and p value = .18 for model 1 and chi = 22.8005 and p value = .1984 in model 2. The poor fit in these models can be explained by a small sample size and a relatively large number of variables. In such cases, research recommends increasing the number of cases, eliminating some predictors, or accepting lessened power for the analysis (Tabachnik and Fidell 2001, 522). For sensitivity analysis, Models 1 and 2 were modified by eliminating all independent variables with the exception of those found to be statistically significant in the full models. Both new models became statistically significant (chi = 11.59 and p value = .009 for model 1; chi = 8.91 and p value = .023 in model 2). All but one independent variable (relationship length in model 2) remained statistically significant in these revised models. Thus, the poor fit of the complete models may be explained by low power, and the findings pertaining to several independent variables shown in models 1 and 2 appear to be theoretically important and statistically significant, despite the overall poor fit of these models.

The first model examines the effect of four groups of antecedents on the likelihood of contractors’ negotiating and discussing performance measurement in an attempt to influence the monitoring process. Three variables show significant results. Nonprofit contractors are more likely to negotiate and discuss performance measures with the government agency. The odds of negotiations pursued by a nonprofit vendor are 3.4 times the odds of negotiations by a for-profit provider. Looking ahead, the nonprofit ownership dummy is insignificant in all other models. Hence, although nonprofit organizations are likely to initiate negotiations, there is no evidence that governments view nonprofit status as a signal of trust and choose to pursue collaborative activities. The reason why nonprofit organizations are more likely to negotiate than their for-profit counterparts may be related to the role governmental performance standards play in these organizations. Profit-seeking organizations are more likely to use financial bottom lines when assessing their own performance.

18 Contact the author for complete results on these models.
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1 Contractor Negotiates and Discusses PM</th>
<th>Model 2 Government Asking Contractors for Input</th>
<th>Model 3 Contractor’s Input is Incorporated</th>
<th>Model 4 Communication Affects PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>(-0.768, 0.464)</td>
<td>(-7.041, 0.001)</td>
<td>(-2.343, 0.096)</td>
<td>(-6.361, 0.002)</td>
</tr>
<tr>
<td>“Hard” services</td>
<td>(0.535, 1.708)</td>
<td>(-1.278, 0.279)</td>
<td>(-2.019, 0.133)</td>
<td>(0.373, 1.452)</td>
</tr>
<tr>
<td>Nonprofit ownership</td>
<td>(1.224, 3.401)</td>
<td>(0.152, 1.164)</td>
<td>(0.448, 1.566)</td>
<td>(0.706, 2.026)</td>
</tr>
<tr>
<td>Contractor’s financial dependency</td>
<td>(0.497, 1.644)</td>
<td>(1.175, 3.237)</td>
<td>(2.534, 12.598)</td>
<td>(1.627, 5.089)</td>
</tr>
<tr>
<td>Contractor’s internal measures</td>
<td>(0.322, 1.380)</td>
<td>(2.043, 7.714)</td>
<td>(0.558, 1.746)</td>
<td>(0.695, 2.004)</td>
</tr>
<tr>
<td>Contractor has unique expertise</td>
<td>(1.230, 3.421)</td>
<td>(1.075, 2.929)</td>
<td>(-0.645, 0.525)</td>
<td>(-0.592, 0.553)</td>
</tr>
<tr>
<td>Contract awarded competitively</td>
<td>(-0.663, 0.515)</td>
<td>(1.678, 5.355)</td>
<td>(-0.676, 0.509)</td>
<td>(1.244, 3.470)</td>
</tr>
<tr>
<td>Environment perceived as dynamic</td>
<td>(-0.100, 0.905)</td>
<td>(-0.793, 0.453)</td>
<td>(-2.770, 0.063)</td>
<td>(-1.823, 0.162)</td>
</tr>
<tr>
<td>Relationship length</td>
<td>(-0.084, 0.920)</td>
<td>(-0.191, 0.826)</td>
<td>(-0.081, 0.922)</td>
<td>(-0.085, 0.918)</td>
</tr>
<tr>
<td>Perceived trust</td>
<td>(-0.629, 0.533)</td>
<td>(0.765, 2.149)</td>
<td>(1.231, 3.425)</td>
<td>(-0.049, 0.952)</td>
</tr>
<tr>
<td>Monitoring: self-reporting</td>
<td>(1.084, 2.955)</td>
<td>(-0.261, 0.771)</td>
<td>(1.699, 5.468)</td>
<td>(2.929, 18.701)</td>
</tr>
<tr>
<td>Monitoring: gov’t inspections</td>
<td>(-0.776, 0.460)</td>
<td>(0.053, 1.055)</td>
<td>(1.180, 3.253)</td>
<td>(0.536, 1.709)</td>
</tr>
<tr>
<td>Monitoring: third party monitoring</td>
<td>(0.749, 2.116)</td>
<td>(1.188, 3.281)</td>
<td>(-1.036, 0.355)</td>
<td>(0.289, 1.335)</td>
</tr>
<tr>
<td>Contractor performance is publicized</td>
<td>(0.100, 1.106)</td>
<td>(-0.918, 0.399)</td>
<td>(0.210, 1.234)</td>
<td>(1.059, 2.883)</td>
</tr>
<tr>
<td>Vendor involvement in PM desirable</td>
<td>(1.288, 3.627)</td>
<td>(0.031, 1.032)</td>
<td>(1.765, 5.843)</td>
<td>(2.598, 13.433)</td>
</tr>
<tr>
<td>Respondent’s work experience</td>
<td>(-0.032, 0.969)</td>
<td>(0.100, 1.105)</td>
<td>(-0.021, 0.980)</td>
<td>(0.067, 1.069)</td>
</tr>
<tr>
<td>Respondent’s contracting experience</td>
<td>(0.000, 1.000)</td>
<td>(-0.015, 0.986)</td>
<td>(0.036, 1.036)</td>
<td>(0.019, 1.019)</td>
</tr>
<tr>
<td>In-house professional capacity</td>
<td>(-1.705, 0.182)</td>
<td>(2.325, 10.229)</td>
<td>(-0.696, 0.499)</td>
<td>(-0.517, 0.596)</td>
</tr>
<tr>
<td>Government/contractor dummy</td>
<td>(0.318, 1.375)</td>
<td>(1.476, 4.375)</td>
<td>(-1.017, 0.362)</td>
<td>(1.000, 2.717)</td>
</tr>
<tr>
<td>Likelihood ratio test (chi and</td>
<td>(23.28, p = .18)</td>
<td>(22.80, p = .19)</td>
<td>(36.67, p = .006)</td>
<td>(31.63, p = .024)</td>
</tr>
<tr>
<td>(p) values shown)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: \(^*\) \(p < 0.1\); \(^{**}\) \(p < 0.05\); \(^{***}\) \(p < 0.01\), Performance Measurement, PM.
One for-profit vendor noted in an interview: “I do evaluate myself to find out what everyone else did, to see whether I am being competitive. My interest is—did I leave money on the table? Did I win big? It is about making money in the end.” Another vendor said: “... if there is a best way of monitoring—it’s through our profit and loss statement every month.” Focusing on bottom lines, for-profits may decide to comply with the process enforced by the government, rather than work on its improvement. On the other hand, measurement strategies used by nonprofit organizations can be similar to those enforced by the government, and hence they will be more likely to integrate the internal and external evaluation systems to avoid administrative burden and redundancy.

Model 1 also suggests that contractors perceived to have a unique expertise or competitive advantage in the local market are more likely to pursue negotiations. This finding confirms the original proposition on the effect of market conditions on the likelihood of pursuing collaborative activities. Unique expertise gives contractors additional leverage in the monitoring process. It is not surprising, therefore, that contractors that are perceived to have this advantage are also likely to attempt to control performance evaluation through negotiation.

Analysis suggests that the greater the professional capacity among the buyers, the lower the likelihood of negotiations initiated by the vendors. This variable also has a significant but positive slope in model 2: the government’s professional capacity increases the likelihood of seeking the contractor’s input in performance monitoring. These findings may suggest that government agencies possessing specialized knowledge are more likely to (a) develop high-quality performance measures accepted across the industry, (b) be trusted by their vendors in the performance evaluation process and hence need less negotiation, and (c) seek vendors’ input, perhaps at the onset of relationships, which would require fewer negotiations later on. This interpretation is corroborated in Model 2: seeking the contractor’s input is more prevalent in long-term relationships. Hence, the contractor’s involvement initiated by the government is more common in the beginning, which provides some evidence of decreased transaction costs in a long-term contractual arrangement.19

The third regression provides evidence of service measurability affecting the use of the contractors’ input. With easily measurable services provided by the “craft” agencies, such as IT or public works, agencies are less likely to incorporate providers’ input, compared to the coping or procedural agencies delivering soft services.20 As suggested in the theory section, governments face greater risk by delegating the delivery of services with low outcome measurability. A decision to incur additional transaction costs by seeking and incorporating vendor input reflects the choice to collaborate when the outcomes are harder to observe.

Although no effect of higher financial dependency on vendor negotiation and input seeking is found, vendor resource dependency predicts incorporating the contractors’ input. The odds of incorporating the vendors’ input in the case of financial dependency is as much as 12 times the odds of doing so with a more financially independent contractor. Vendors, whose fiscal health depends largely on one contract, may find alternative, implicit ways of influencing the evaluation process. Overall, these findings confirm that resource dependency results in the vendors’ participation in performance evaluation.

---

19 This variable became insignificant in the sensitivity analysis described above.
20 No organization in the sample could be classified as a “production agency.”
Incorporating the vendors’ input and changing measurement practices is also less likely to occur in the environments perceived as somewhat or highly dynamic. Rapid changes in service needs, technology, competition, and funding discourage government agencies from engaging in dialogue with their vendors and making modifications in the evaluation process. In such cases, contracts are not perceived as stable and long term, and hence collaboration is not expected to have a “reasonable payoff” (Milward and Provan 2000, 376).

Regressions suggest that change in performance measurement as a result of communicating with the vendor is significantly and positively associated with the use of self-reported data. When data collection and reporting tasks are delegated to the vendor, the latter directly participates in the performance evaluation process, and, therefore, is more likely to identify problems, seek coherence between external and internal evaluation, and hence, attempt to influence these procedures.

Finally, there is some evidence of collaboration occurring when participants perceive the contractors’ role as important in the performance evaluation process. Respondents, who agree that contractors should be engaged in the design of monitoring systems, reported that their systems were in fact affected by the vendors’ input. Here, the possibility of a recursive relationship cannot be eliminated: subjects who practice collaborative performance monitoring may be more likely to support this practice. Our use of cross-sectional data does not allow us to rule out endogeneity of the respondent’s support of collaborative practices.

Several predictors included in the theoretical model failed to produce significant results. Thus, the contractors’ use of internal performance evaluation techniques does not appear to enhance or disrupt the dialogue between government agencies and their vendors. Across all models, contracts awarded competitively are no more likely to stimulate collaboration than those awarded without competitive bidding. Controlling for the government’s professional capacity, the degree of trust reported by respondents is not associated with the likelihood of negotiations and input seeking. Regression analysis also suggests that approaches taken to collect performance information do not influence the government’s or the vendor’s ability to work together. Finally, most respondent characteristics appear to be insignificant: the respondents’ work and contract management experience, as well as being a monitoring officer rather than a vendor are not associated with collaborative activities.

Is the analysis presented in this study sufficient to revise the model of collaboration determinants shown in figure 2? In addition to the limitations pertaining to generalizability, one important consideration related to the findings of this study precludes us from a positive answer to this question: although some overlaps exist, clearly, a different set of determinants played a role in the prevalence of each type of collaboration used in this study. Thus, with the identified richness of collaborative activities, one might question the validity of a single overarching model predicting collaborative behavior in general. Rather than proposing a revised model, this article suggests that several factors determine certain collaborative activities, and these findings, along with the typology of collaborative activities in a variety of contexts, should be explored in future research.

CONCLUSION

The fundamental issue of American government is related to the process of drawing organizational boundaries (Kettl 2006). Recognizing the arbitrariness of this decision, some managers build new boundaries, whereas others enhance horizontal collaboration and
develop administrative structures that involve multiple decision makers (Kettl 2006). This study is an attempt to examine how contract officers in state and local governments draw boundaries while creating performance measurement systems. A key task of contract monitoring involves aligning goals, building trust, and conducting evaluations that are not formalistic but complex and tailored to specific contracts. The objective of this research is to examine the extent to which government agencies and their vendors share power while designing such performance systems.

The findings of this study have several theoretical implications for the way researchers approach contracting relationships. First, this study provides new evidence suggesting that the view of contracting dyads as merely buyers and sellers, or principals and agents, is insufficient. In this study, a large share of contracts incorporated a variety of joint activities aimed at adjusting and improving performance evaluation. Acknowledging their unique expertise and financial interests, vendors actively internalize government performance evaluation procedures and participate in defining the expectations and evaluation techniques. These findings support the application of stewardship theory, as originally proposed by Van Slyke (2007). The theory stresses the role of collective behavior in which goal convergence rather than self-interest is prioritized, and vendors are involved in goal formulation and decision making.

Second, contracting discourse often dichotomizes relational contracts, based on reciprocity, discretion, and few checks and balances, and transactional contracts, relying on compliance with prespecified standards, rather than trust or collaboration. This study questions the utility of such dichotomy. Collaboration and formal performance measurement do not appear to be a zero-sum game. Instead, they coexist, and joint decision making may in fact validate and enhance the existing contract monitoring procedures.

Finally, this study supports the idea that collaboration is more than just an institutional form, but rather an individual-level activity. Contracts, networks, cross-sector and cross-jurisdictional partnerships may incorporate a different degree of joint decision making as well as some degree of opposition toward collaboration. Many collaborative activities examined in this article were in fact informal. This has important methodological implications for studying collaborations: they are more appropriately studied through in-depth interviews, combined with document analysis, observations, and other qualitative research methods.

This article also contributes to the limited empirical literature on contract monitoring and performance measurement. Respondents in this study utilize a wide variety of monitoring techniques and focus on different aspects of performance. Nonetheless, confirming the findings of past research (Van Slyke 2003), monitoring officers’ concerns about their capacity to effectively monitor performance still persist. Respondents report lacking the skills and the resources needed to design the tools, collect appropriate data, and enforce performance standards. Notably, evaluation of cost effectiveness and service impact on citizens, which some may view as essential in performance evaluation, is lacking compared to other measures examined in this study. In light of the traditional privatization promise to cut costs and improve quality, this study demonstrates that little is done to verify these claims and to compare contracting to in-house service delivery. Furthermore, although the general pattern of performance monitoring was similar across the samples of monitoring officers and vendors, the latter group reported less monitoring than their public counterparts. Some vendors were not aware of being evaluated and had little knowledge of how
performance information was used. If performance measurement is intended as a preventative rather than a punitive measure, expected to reduce moral hazard, vendors’ ignorance of government monitoring may reduce its effectiveness. These findings suggest that governments should educate their vendors about the performance measurement process especially if monitoring is intended as prevention and enforcement, rather than merely an evaluation tool.

Contractors in our sample develop and propose new measures; they point out the limitations of the existing performance standards and negotiate the adjustment of these systems to their respective contexts and clientele. Such assertive participation in the performance evaluation process may present both challenges and opportunities for effective contract implementation. On one hand, the contractors’ input can help gain better understanding of issues arising in the course of service delivery and a broader perspective of service outcomes. The use of collaborative strategies promotes data sharing and provides opportunities for catching the problems early on. Collaboration creates a positive and open culture and helps avoid generic, formalistic, and unreasonable procedures. It can also motivate and empower contractors by giving them a chance to share their input. On the other hand, the vendors’ participation necessitates an evaluation of their claims. This may not always be feasible: an agency lacking the capacity to develop effective monitoring tools may also lack the capacity to assess the vendors’ input. There are also ethical concerns associated with treating vendors as coequals in this process. The latter assumes partial replacement of traditional accountability mechanisms with self-responsibility, professionalism, trust, and collegiality of participating agents (Agranoff and McGuire 2001). Since performance evaluation is more relevant to the “governance” rather than the “service delivery” role of public agencies, involving providers raises questions about democratic accountability, creates threats to objectivity, and perhaps limits the government’s ability to be a smart buyer (Van Slyke 2003, 306). Cooperative relationships may be perceived as collusion and cooptation and may become more important than performance by creating bias towards some contractors (DeHoog 1990).

This complex environment has important implications for practitioners in the field of contracting, as noted by a city manager, who served as a discussant for this manuscript at the 2008 Association for Public Policy Analysis and Management annual conference. He stressed that contract officers commonly operate in the bottom layers of organizational hierarchies. Despite the level of discretion they are granted, their skills and training are rarely at the center of attention of the top leadership. Meanwhile, their role must be seen as essential to effective implementation of government programs and efficient use of public tax dollars.

Thus, government executives should acknowledge that the monitoring officers’ role in a collaborative setting goes beyond specifying and enforcing a set of performance standards. Contract implementation may involve a complex decision-making process requiring a skill set more in line with collaborative rather than hierarchical principles of management. It requires the ability to empower and encourage contractors to share their perspectives on service outcomes; willingness to learn from each other; and professional knowledge to understand the context and evaluate the partners’ claims. High prevalence of instances warranting some change in performance monitoring also reflects the need to build flexibility into the contracts in order to encourage all parties to make changes when necessary. These skills should be continuously evaluated and enhanced in the public sector. The analysis of
collaboration determinants presented in this study provides some evidence of collaborative management capacity. Government agencies attempt to minimize their transaction costs by reducing collaboration in cases of high service measurability, long-term relationships, and highly dynamic markets. Governments with more professional capacity also face less negotiation and effectively rely on the contractors’ input while developing performance management systems.

Importantly, collaboration is only a means to an end: when properly applied it can generate public value and help make a difference (Entwistle and Martin 2005). A question that remains unanswered in this study pertains to the effect of collaborative practices on the quality of performance measures and the overall effectiveness of contract implementation. Future research should determine whether collaborative practices result in a more diverse and tailored set of monitoring tools. Another question warranting exploration pertains to the effect of vendor ownership on (a) the monitoring process, (b) collaboration, and (c) vendor outcomes. Understanding the role of ownership will help design performance monitoring tools utilizing the strengths of each sector. Finally, the results of this study should be verified using larger samples that include paired data on monitoring officers and their vendors. Descriptive statistics analysed here presume that public officers may perceive monitoring and collaboration processes differently from their private counterparts. Analysis of contracting dyads would allow a more in-depth examination of differences in these perceptions.

APPENDIX 1

Select Interview Questions Used to Examine the Prevalence and the Process of Collaboration

1. In some cases government agencies collect and monitor all the information pertaining to a contractor’s performance directly. In other cases, governments use information collected and provided by the contractor (so-called self-reported measures). There are also agencies that use third parties to collect information and do the monitoring (for instance the clients or 3rd party inspectors). Which strategy do you use?  

2. Today I would like us to talk about performance evaluation and measurement, specifically, about any kind of information that you might use to make sure that your contractor is complying with your expectations and doing their job well. Some of these performance measures can be more formal and quantitative (e.g. reporting the number of service units produced every week). Other measures can be more informal (e.g. informally discussing service provision details). For the following questions please choose one of the following answers: (1) yes, (2) no, (3) don’t know/don’t recall/refuse to answer. In this contract, do you collect, monitor, or evaluate information on:
   a. cost-effectiveness of contracted services?
   b. quality of services provided by the contractor?

21 Questions listed in Appendix 1 were used in the interviews with government employees. Conceptually identical questions were used in vendor interviews, although wording has been changed (e.g. the question “Do you evaluate or monitor your contractor’s performance?” was modified to “Does the government agency evaluate or monitor the performance of your organization?”).
c. contractor’s workload (e.g. number of clients served, units of services provided, # hours of work)?
d. the impact that services have on clients or service-recipients?
e. customer satisfaction?
f. contractors’ ability to provide equitable access to services without any discrimination (e.g. based on income, gender, or race)?
g. compliance of service provision with the law?
h. timeliness of service delivery?
i. service continuity or any disruptions in service delivery?
j. do you specify the detailed procedures for service delivery; in other words, precisely how services should be delivered, and by whom?
k. do you use any quantitative measures of performance (for instance number of clients served, # services provided, quantifiable impact on the clients’ status)?
l. do you use any qualitative, descriptive information on your contractor’s performance?
m. do you use any informal ways of obtaining performance information, such as through an informal conversation with a client, contractor staff or a third party?
n. are the performance measures that you use tailored to this particular contractor (i.e. they wouldn’t be used for another contractor)?
o. do you collect information on the reputation of your contractor formally or informally?
p. do you evaluate your contractor’s performance in any other ways that I have not listed?

3. Do you evaluate or monitor your contractor’s performance? If yes, ask: In what way?

4. Have you ever asked for contractor’s input on the performance measures that are used in this contract? If yes, ask: Why was this done? Can you explain how this was done?

5. Has your contractor ever attempted to negotiate or discuss with you how their performance should be measured or evaluated? If yes, ask: When did this happen? When this happened, how did your agency respond?

6. Have you incorporated any performance measures that were modified in response to contractor’s comments or proposed by your contractor?

7. Do you believe that your overall communication with the contractor influenced the type of measures that you use? If yes, ask: In what way?
APPENDIX 2

Table A1
Measurement of Control Variables in the Regression Analysis

<table>
<thead>
<tr>
<th>Interview Questions Used to Create Each Variable</th>
<th>Collaboration Determinants and Created Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service measurability</strong></td>
<td><strong>Collaboration Determinants and Created Measures</strong></td>
</tr>
<tr>
<td>Responses were categorized into “hard” (easily measurable) services: IT, construction, maintenance, public works, planting and plant control, food supply and quality monitoring, animal care, janitorial, translation, and recreational (camps, dance lessons). “Soft” (hard-to-measure) services: long-term care, medical, nursing care, health management, mental health, psychological consultation, arts therapy, programs for women and children, consulting, evaluation and training, criminal justice, animal care, substance abuse, and homelessness (dummy variable). Using Wilson’s classification, hard services also corresponded to those provided by coping and procedural agencies, whereas soft ones corresponded to so-called craft agencies (no production agencies were involved in the analysis).</td>
<td></td>
</tr>
<tr>
<td><strong>Contractor ownership</strong></td>
<td><strong>Contractor financial dependency on the contract</strong></td>
</tr>
<tr>
<td>Variable “nonprofit” was created to indicate nonprofit status of the vendor in the discussed contracting arrangement (dummy variable).</td>
<td></td>
</tr>
<tr>
<td><strong>Contractor financial dependency on the contract</strong></td>
<td></td>
</tr>
<tr>
<td>Variable “Contractor’s financial dependency” was coded as one for responses (a) financially very dependent on government funding and (b) somewhat dependent, and 0 for all other options (dummy variable).</td>
<td></td>
</tr>
<tr>
<td><strong>Contractor uses internal performance measures</strong></td>
<td></td>
</tr>
<tr>
<td>Variable “internal measures” was created and coded as one for affirmative responses to the question (dummy variable).</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Interview Questions Used to Create Each Variable</th>
<th>Collaboration Determinants and Created Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does your contractor have a unique expertise that is difficult to find elsewhere? (probe: are there any other organizations in this area that provide similar services? Is this market very competitive?)</td>
<td>Market competition (multiple suppliers with comparable expertise exist)</td>
</tr>
<tr>
<td>Did you go through the process of competitive bidding for this contract?</td>
<td>Variable “unique expertise” was created for affirmative responses (dummy variable).</td>
</tr>
<tr>
<td>Some contracts exist in the fields that undergo rapid changes in service needs, service technology, suppliers, or funding. Other contracts exist in more stable, less uncertain environments. Where would you place this contract on this continuum between very dynamic and very stable environments? (a) Very dynamic, (b) somewhat dynamic, (c) somewhat stable, (d) very stable.</td>
<td>Variable “competitive bidding” was coded as one for affirmative responses (dummy variable).</td>
</tr>
<tr>
<td>When was this contract initiated? Have you been working with this contractor before? If yes, ask: In what capacity? For how long?</td>
<td>Dynamic versus stable environment</td>
</tr>
<tr>
<td></td>
<td>Variable “dynamic environment” was coded as one for responses (a) very dynamic and (b) somewhat dynamic (dummy variable).</td>
</tr>
<tr>
<td>Some practitioners say that contractual relationships often begin as more rigid (more formal) and over time evolve into relationships that are based on trust. Has this been the case with this contract?</td>
<td>Long-term versus short-term relationship</td>
</tr>
<tr>
<td></td>
<td>Variable “relationship length” was created reflecting the number of years (interval-ratio variable). Eight cases in the data had missing values and median values were imputed in order to retain this variable and maximize the size of the sample.</td>
</tr>
<tr>
<td>In some cases, government agencies collect and monitor all the information pertaining to contractor’s performance directly. In other cases, governments use information collected and provided by the contractor (so-called self-reported measures). There are also agencies that use third parties to collect information and do the monitoring (for instance, the clients or third party inspectors). Which strategy do you use?</td>
<td>Perceived goal congruence or trust</td>
</tr>
<tr>
<td>Do you publicize the information pertaining to the performance of the contractor? If yes, ask: In what way?</td>
<td>Variable “trust” was coded as one for responses confirming that the relationship was presently characterized by trust between the government agency and the vendor (dummy variable).</td>
</tr>
<tr>
<td></td>
<td>Monitoring: Self-reported measures used; Monitoring: government collects performance information; Monitoring: third party monitoring is used</td>
</tr>
<tr>
<td></td>
<td>Three dummy variables (“self-reporting”, “government inspection”, “other monitors”) were created based on the descriptive answer to the question.</td>
</tr>
<tr>
<td></td>
<td>As an additional measure of third-party monitoring, variable “publicized performance” was created and affirmative responses to the second question were coded as one (dummy variable).</td>
</tr>
</tbody>
</table>

Continued
### Table A1 (continued)
**Measurement of Control Variables in the Regression Analysis**

<table>
<thead>
<tr>
<th>Interview Questions Used to Create Each Variable</th>
<th>Collaboration Determinants and Created Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think contractors should be engaged in the development of performance measures to oversee their own work? If yes or no, ask: can you explain, why?</td>
<td><strong>Contractor’s participation in performance evaluation is perceived as desirable</strong></td>
</tr>
<tr>
<td>Note for the interviewer: verify respondent’s employment status. Do you currently serve as ______________? How long have you been working in this position? How long have you been involved in managing contracts?</td>
<td><strong>Respondent’s work and contract management experience</strong></td>
</tr>
<tr>
<td>Do you have professionals among your staff who can thoroughly understand the nature of the service delivered by your contractor (individuals with similar education, degrees, professional norms, etc.)?</td>
<td><strong>In-house capacity to deliver the service</strong></td>
</tr>
</tbody>
</table>

---

**REFERENCES**


