

## ABSTRACT

Title of Document: THE PERFORMANCE OF PERFORMANCE-BASED CONTRACTING IN HUMAN SERVICES

Jiahuan Lu, PhD, 2014

Directed By: Dr. Donald F. Kettl, School of Public Policy

Performance-based contracting (PBC) is becoming increasingly attractive to public human service agencies. By attaching contract compensation to contractors' performance achievement, PBC is expected to encourage quality services, better outcomes, and less administrative monitoring. However, the burgeoning popularity of PBC lacks sufficient evidence to confirm these promised benefits. In particular, the efforts of introducing PBC into human service systems needs first to address the effectiveness problem, i.e., whether PBC really produces better results. This problem constitutes the research question of the research project.

After building the theoretical framework which incorporates the literature on formal and relational contracting, this project explores the effectiveness question using Indiana vocational rehabilitation program as a case. In particular, the study evaluates PBC effectiveness from two perspectives: service outcome and participating organizations. From a service-outcome perspective, the research employs a quasi-

experimental design to compare the impacts of two contract arrangements, PBC and fee-for-service (FFS), on individual employment outcomes. From a participating-organization perspective, the project runs semi-structured interviews with service counselors and contractors. Triangulating these findings, this project proposes that PBC seems more promising than FFS in human services. It also implies PBC effectiveness might not be well-rounded and should not be exaggerated.

Further, the study addresses the managerial implications of the findings. The research and the practice of PBC tend to ignore the relational face of contracting. PBC as a formal arrangement is always disturbed by the highly uncertain nature of human services and thus might result in incomplete performance improvement and contractor opportunism. If so, relational contracting, using informal and normative mechanisms, may enable desirable collaborative outcomes. The combination of formal PBC efforts with relational contracting would encourage high-quality results.

In sum, this project represents an attempt to systematically examine PBC effectiveness in human services. It shows the difficulties and dynamics of introducing performance management to human service contracting. It also warns the launch of PBC systems should be very deliberate and careful. More broadly, the project underscores two key components of contracting management: control and trust.

THE PERFORMANCE OF PERFORMANCE-BASED CONTRACTING IN  
HUMAN SERVICES

By

Jiahuan Lu

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2014

Advisory Committee:  
Professor Donald F. Kettl, Chair  
Professor Philip Joyce  
Professor Steven Rathgeb Smith  
Professor Jocelyn Johnston  
Professor Ellen Fabian

© Copyright by  
Jiahuan Lu  
2014

## Dedication

To those who helped me complete this dissertation.

## Acknowledgements

To me, public administration is so fascinating and exciting that I feel lucky to be part of it. The dissertation marks the end of my doctoral study. This moment has certainly prompted a lot of reflection on my part on my intellectual and personal journey in the last five years. Over the years, the intellectual debts I have accumulated are very large.

I owe a special debt to my advisor, Dr. Don Kettl. Dr. Kettl and I came to the University of Maryland both in 2009. Since then, I have benefited greatly from numerous discussions with him. I know it would be very difficult for a dean to work very closely with doctoral students over five years, but Dr. Kettl did that. He is always insightful and supportive, answering my questions on public administration and fundamentally shaping my understanding of public administration.

Dr. Steve Smith is always nice and patient. We had a conversation at Georgetown University in 2011 and Dr. Smith offered to serve on my committee. That conversation helped narrow down my research questions on performance-based service contracting and further led to the dissertation project presented here. Since then, we met periodically and I was fortunate to share his thoughts on nonprofit management and service contracting.

I am grateful indeed to the Department of Public Administration and Policy at American University for the support for my coursework. Dr. Bob Durant and Dr.

Jocelyn Johnston allowed me to take two core doctoral-level courses, Proseminar in Public Administration and Seminar in Public Management, which deeply formed my knowledge base of public administration and management. Dr. Johnston later became a member of my dissertation committee and I thus have a chance to access her expertise on social policy and government contracting.

I would also want to thank Dr. Phil Joyce and Dr. Ellen Fabian. Dr. Joyce provided excellent advices on structuring and editing the dissertation. Dr. Fabian was very nice to serve on my committee as the Dean's Representative and provided much guidance on vocational rehabilitation from both theoretical and practical perspectives.

In addition, throughout my dissertation process, there are so many people who extended their help to me but I could not mention in detail here. For example, program managers, area supervisors, and service counselors in both Oklahoma and Indiana vocational rehabilitation agencies, such as Teri Egner, Theresa Koleszar, and Kristy Cook, were incredibly open to me. We made phone calls and they were willing to share their experience on performance-based contracting without reservations.

Again, scholars from other fields such as John McGrew, Grant Revell, and Lawrence Martin kindly provided insights on performance-based contracting from their own perspectives at the very beginning of this project. All these supports make this dissertation possible and urge me to continue my exploration in my fields of interest.

# Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
List of Tables .....	vi
List of Figures.....	vii
Chapter 1. The Rise of Performance-based Contracting .....	1
1.1 The Management Imperative under the Contracting Regime.....	1
1.2 Performance-based Contracting as a New Experiment .....	7
Chapter 2. Burgeoning Popularity, Different Models, and Elusive Effectiveness .....	15
2.1 The Evolution of PBC at Federal Level.....	15
2.2 Popularity at State and Local Levels .....	18
2.3 Important but Missing Links.....	28
Chapter 3. Theoretical Framework .....	41
3.1 Formal Contract Design: A Principal-agent Perspective .....	41
3.2 Formal Contract Design for Human Services.....	45
3.3 Informal Contract Design: A Relational Contracting Perspective.....	52
Chapter 4. Vocational Rehabilitation as a Policy Field .....	61
4.1 Vocational Rehabilitation Programs .....	61
4.2 The Purchase of Job-related Services .....	63
4.3 PBC Models in VR Services.....	69
4.4 The Use of PBC in VR Services.....	78
Chapter 5. The Effectiveness of PBC: A Service Outcome Perspective .....	82
5.1 Introduction.....	82
5.2 Research Design.....	85
5.3 Interrupted Time Series with a Control Group Design.....	87
5.4 Propensity Score Matching .....	93
5.5 Difference-in-difference Regressions .....	103
5.6 Conclusion .....	110
Chapter 6. The Effectiveness of PBC: Government and Contractor Perspectives ...	115
6.1 Introduction.....	115
6.2 Street-level Perspective in Policy Analysis .....	115
6.3 Vocational Rehabilitation Context and Data Collection.....	121
6.4 Findings from VR Agency Perspective .....	122
6.5 Findings from Contractor Perspective .....	126
6.6 Conclusion .....	129
Chapter 7. Two Faces of Contracting, Two Kinds of Control.....	133
7.1 The Effectiveness of PBC as a Formal Arrangement .....	133
7.2 Managerial Implications from a Relational Contracting Perspective .....	137
7.3 Conclusion: Control, Trust, and Contracting Management .....	146
Bibliography .....	151

## List of Tables

Table 1. A Brief History of PBC in the Federal Government .....	17
Table 2. PBC in Maine Substance Abuse Treatment Services .....	20-21
Table 3. PBC in Delaware Alcohol and Other Drug Treatment Programs .....	22
Table 4. State Use of PBCs in 2009 .....	23-24
Table 5. PBC in Selected State Child Welfare Agencies .....	24-25
Table 6. Selected Studies on PBC Effectiveness in Human Services .....	33-35
Table 7. The Determinants of Organizational Control Strategies .....	43
Table 8. Contract Type for Human Service Contracting .....	52
Table 9. Comparisons between Formal and Relational Contracting .....	54-55
Table 10. Components of VR Services and Contract Type .....	67-68
Table 11. Oklahoma Milestone Payment System .....	71-73
Table 12. New York State Milestone Payment System .....	75
Table 13. Indiana Result-based Funding System .....	77-78
Table 14. Description of Matching Variables .....	94-95
Table 15. Covariate Balance Check Before and After Matching .....	101-102
Table 16. Logistic Regression Model Predicting Likelihood of Employment for Service Recipients .....	105-106
Table 17. OLS Regression Models Analyzing Employment Outcomes .....	107-108
Table 18. Distribution of Interview Samples .....	122
Table 19. Mode of Trust Production and Implications for PBC .....	141

## List of Figures

Figure 1. A System Framework of Service Process .....	10
Figure 2. Performance Reform Hierarchy .....	13
Figure 3. The Determinants of Contract Type .....	44
Figure 4. The Contractual Relationship in the Purchase of Job-Related Services ....	64
Figure 5. Analytical Framework for Network Effectiveness .....	83
Figure 6. Interrupted Time Series with a Nonequivalent Control Group Design ....	88

## Chapter 1. The Rise of Performance-based Contracting

### 1.1 The Management Imperative under the Contracting Regime

In *Preface to Public Administration*, Stillman (1991) delineates the “stateless” origins of American public administration. He argues a systematic design or thought of public administration was absent at the founding of the United States. “America’s ‘missing state’ at its inception,” Stillman (1991) believes, “fundamentally shapes our way of thinking about, as well as doing, public administration today” (19). Further, to prevent abuse of public power, private power was relied on as much as possible. In this line, America has an ingrained tradition of using private power, in addition to public administrative capacity, to solve public problems (Kingdon, 1999).

Government contracting, the most common type of privatization (Savas, 1987), is thus so widely and durably used as a government tool that it has become a remarkable feature of the American governance system. At all levels of governments, the massive use of government contracting to provide public goods and services and achieve policy priorities is a common government practice. Accordingly, “government by proxy” (Kettl, 1988), “hollow state” (Milward & Provan, 2000), and many other labels have been attached to the public administration narrative.

In recent years, government contracting is becoming more dynamic, including not only contracting out, but contracting back-in. Decades of contracting-out experiences have rationalized governments’ contracting-out decisions. The insufficiency of

contracting management and monitoring capacities has given rise to “reverse contracting”, restoring from third-party delivery back to in-house delivery (Hefetz & Warner, 2004). Even so, there is still no evidence to see the ebb of contracting. Virtually every government is dependent on contracts to a varying degree. At the federal level, according to [USASpending.gov](http://USASpending.gov), more than one third of federal spending is used under the titles of “Contracts” and “Grants” from fiscal years 2000 to 2010 on average<sup>1</sup>. Again, at the local level, the scope of contracting is equally prominent. Approximately 45.5% of local government services are delivered through contracting in 2007 (Warner & Hefetz, 2009). A recent survey of U.S. local government managers shows that 93% of municipal officials support government contracting (Girth & Johnston, 2011).

The field of human services is an indispensable component of government contracting. In fact, the use of government contracting in human services occurs much earlier than contracting for other goods and services in the United States. The historic roots can be dated back to the colonial period (Smith & Lipsky, 1993). Today, governments at every level do very little directly by themselves in human service provision. Rather, they fund third-party actors through government contracts to provide services (Salamon, 1995). Among them, nonprofits deliver a large share of government funded human services. All over the United States, for example, 56.3% of homeless shelters, 35.9% of drug and alcohol treatment programs, and 32.8% of day care facilities are run by nonprofits in local communities (Warner & Hefetz, 2009). In 2009, governments at all levels contracted with 33,000 human service

---

<sup>1</sup> Data come from [www.usaspending.gov](http://www.usaspending.gov) (accessed on January 30, 2012).

nonprofits for approximately 200,000 contracts and grants worth over \$100 billion (Boris, de Leon, Roeger, & Nikolova, 2010). Such an extensive government-nonprofit partnership features the U.S. human service delivery system, termed by Smith and Lipsky (1993) as “contracting regime” or by Salamon (Salamon, 1987) as “third-party government”.

The significant explosion of contracting fundamentally reshapes the features and businesses of government. In a political sense, contractors constitute an important pillar of American institutions in serving democratic governance and citizenship (Smith & Lipsky, 1993, Cooper, 2003). The policy goals and missions of grand federal and state programs now depends on contractors to represent and realize. Contractors thus act as a critical buffer between the state and citizens. In a managerial sense, since government programs are dependent on contract operation, the performance of government turns to be largely contingent on contractors (Frederickson & Frederickson, 2006; Kettl, 2002). In short, sound contracting performance will not only directly improve government performance, but indirectly improve democratic governance (Behn, 2002). This further raises the critical issue of contracting management. “It makes no sense to speak of effective public policy or of professional public management, or even informed citizenship, without an awareness of the nature and operation of public contract management” (Cooper, 2003, 12).

Indeed, contracting is not a panacea with self-enforcing nature. Government’s retreat from human service delivery and reliance on contract operation, no matter aiming for

efficiency improvement, “load-shedding,” or both, does not simply eliminate government role. Although contractors provide various services to citizens as proxies of the state, government continues to bear the responsibility for satisfactory service delivery. The explosion of contracting actually calls for a different government role. Using Osborne and Gaebler’s (1992) metaphor, government now should be “**steering not rowing.**” The governance by contracting mode gives prominence to contracting management capacity, i.e., if government is able to act as a “smart buyer” throughout the contracting process (Kettl, 1993). “The most fundamental problem with the current system,” as Kelman (2002) suggests, “is that it insufficiently recognizes contract administration as in the first instance a management function” (93). Given the large scope of contracting employed by governments, he further argues that “the ability to manage contracting must be considered a core competency of the organization” (89).

However, managing indirect government tools is very different from managing goods and service production within traditional government bureaucries. A central puzzle for public managers in contracting management, as Kettl (2002) summarizes, is that “[t]hey are responsible for ensuring high-quality results in programs that they do not directly control” (493). The reliance on contracting in public management represents a significant shift away from a vertical, authority-based model to a horizontal, negotiation-driven model (Cooper, 2003). When government directly delivers services, there is a clear chain of commands within government domain and all the managerial behaviors are based on hierarchical authority. However, when various

indirect government tools are introduced into governance system, such authority relationship is absent. All the relationships underlying indirect government tools are now based on voluntary market exchange. “The basic administrative problem of indirect government thus is developing effective management mechanisms [such as bargaining and incentive system] to replace command and control” (Kettl, 2002, 491). Therefore, the managerial responsibility turns “to arrange networks rather than to carry out the traditional task of government, which is to manage hierarchies” (Milward & Provan, 2000, 362).

Effective contracting management requires public managers’ sensible answers to the questions of “what to buy, who to buy it from, and what it has bought” (Kettl, 1993, 180). Accordingly, it calls for “personnel with contract-management experience, policy expertise, negotiation, bargaining, and mediation skills, oversight and program audit capabilities, and the necessary communication and political skills to manage programs with third parties in a complex political environment” (Van Slyke, 2003). However, in contrast to the ubiquitous use of contracting and the critical role of contracting management within is the finding that contracting management capacity is often insufficient, which “create[s] serious public management and accountability problems for which public administration theory fails to prepare us” (Salamon, 1989, 11). For example, Van Slyke (2003) finds serious capacity shortage in social service contracting management in New York state, as demonstrated by loss of contract management expertise, institutional memory, and capacity constraints. Smith and Smyth’s (1996) study of substance abuse service contracting in North Carolina shows

that limited administrative resources (personnel and budget) undermine contracting management capacity, and program evaluations are often difficult. When examining contracting management in local governments from 1997 to 2007, Joaquin and Greitens (2012) observe a significant decline in management capacity in agenda setting, formulation, and implementation. This decline is even more significant as local governments contract out more complex goods and services.

“The poor management of service contracts,” the U.S. Government Accountability Office (GAO) (2001) concludes, “undermines the government’s ability to obtain good value for the money spent” (5). The deficit of management capacity in the use of contracting would incur substantial uncertainty in aligning private market with public interest (Kettl, 1993). “Any uncertainty surrounding the relation between market means and public ends, any range of discretion or ambiguity,” as Donahue and Nye (2002) argues, “will result, ... in effort gravitating toward the focus of intensity (private interest)” (7-8).

In short, contracting management is a demanding and distinct craft. To address this challenge, public management scholarship in the last two decades was marked by a surge of exploration of with various capacity-building mechanisms, such as rationalizing make-or-buy decisions by balancing contracting out and contracting back-in (Brudney, Fernandez, Ryu, & Wright, 2005; Hefetz & Warner, 2004; Johnston & Romzek, 1999), managing thin service markets to stimulate and maintain competition (Brown & Potoski, 2004; Johnston & Girth, 2012; Warner & Hefetz,

2008), relying more on relational contracting to supplement formal contracts (Bertelli & Smith, 2010; Van Slyke, 2007), designing appropriate monitoring tools to tailor contractor incentives and ownership (Amirkhanyan, 2010; Lambright, 2009), improving contract design (Kim & Brown, 2012), and discovering new accountability mechanisms (Romzek & Johnston, 2005; Romzek, LeRoux, & Blackmar, 2012). Under this context, performance-based contracting (PBC), incorporating performance incentives into contract specification and compensation, comes to the agenda of public managers at all levels of governments.

## 1.2 Performance-based Contracting as a New Experiment

Currently, performance-based contracting (PBC) is enjoying a widespread popularity and acclaim as a preferred contracting approach in government acquisition of a variety of goods and services. PBC may also be referred as to result-based contracting, performance-based acquisition, and result-based funding in different contexts. Despite the burgeoning popularity, the connotation of PBC is still quite elusive. Martin (1999) thinks PBC “focuses on the outputs, quality and outcomes of service provision and may tie at least a portion of a contractor’s payment as well as any contract extension or renewal to their achievement” (1). Cooper (2003) considers PBC should “include incentive and penalty clauses that provide benchmarks to assess performance as well as mechanisms to encourage contractors to exceed those minimum levels and to do so at a lower cost than that absolutely required under the contract” (98).

Federal acquisition regulation (FAR) provides a more technical definition:

“Performance-based contracts (a) describe the requirements in terms of results required rather than the methods of performance of the work; (b) use measurable performance standards (i.e., terms of quality, timeliness, quantity, etc.) and quality assurance surveillance plans; (c) specify procedures for reductions of fee or for reductions to the price of a fixed-price contract when services are not performed or do not meet contract requirements; and (d) include performance incentives where appropriate” (FAR 37.601). In this way, PBC could “ensure that required performance quality levels are achieved and that total payment is related to the degree that services performed meet contract standards” (FAR 37.601).

In this study, PBC is defined in a loose way as an “umbrella” term: PBC incorporates performance measures in contract specifications and makes contract compensations (such as payment, extension, and renewal) fully or partially contingent on performance achievements. Suggested by FAR 37, a performance-based contract may include (1) a performance-based work statement, which specifies the work in a quantifiable and measurable way; (2) measurable performance standards in terms of quality, quantity, and timeliness; (3) methods measuring contractor’s performance against the performance standards; (4) performance incentives tied to the performance standards (FAR 37.6). Based on this, public agencies have designed a variety of PBC models. The variance between different models largely centers on five dimensions: (1) payment schedule, (2) the extent to which incentives/disincentives are used, (3) frequency of performance reporting, (4) the extent of which providers are involved in

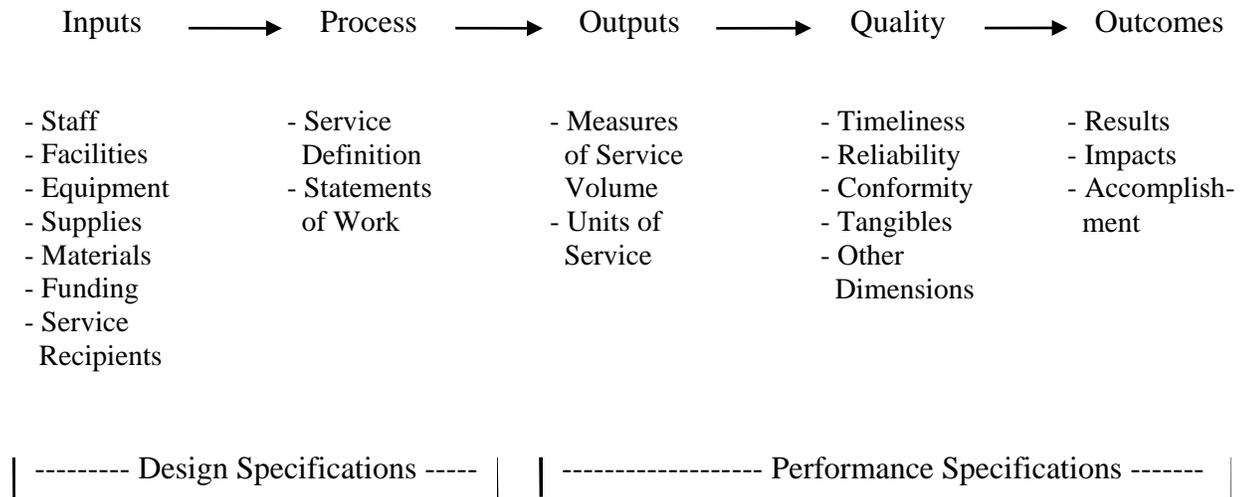
performance indicator development, and (5) level of financial risks assumed by contractors.

Traditionally, services are procured using a fee-for-service (FFS) contracting approach. This contracting method specifies the standards on inputs and delivery process, such as amount of time and labor required, and detailed procedures to be followed in delivering services. When services are delivered, contractors are reimbursed based on units of service delivered throughout the service process. Compared with FFS, conceptually, PBC represents several substantial changes in the landscape of service contracting.

First, PBC changes contract specification method, from a design specification (focusing on input and process) to a performance specification (focusing on output, quality, and outcome) (See Figure 1) (Martin, 2005). Under PBC, public managers clearly specify the desired end results that service contractors should achieve, while leaving contractors considerable flexibility and freedom to prescribe service methods and use of funds to accomplish those goals. By tying contract compensation contingent on performance achievement, public managers contract for service outcomes, no longer for services per se. Relating to the change in contract specification, PBC presents new challenges for public managers. Under FFS, public managers are responsible for specifying the standards on inputs and the details of service process, ensuring the delivery of promised services. Using PBC, public

managers are expected to specify outcomes, designing incentives, and evaluating outcomes, leaving contractors to produce desired results.

Figure 1. A System Framework of Service Process



Source: Martin (2005).

Second, the same as other performance management strategies, PBC implies a change in accountability mechanisms, with increasing attention to accountability for results. This, using Al Gore's (1993) words, represents "a fundamental shift in the system of accountability ... from one oriented around accountability for processes and inputs to one that measures performance and is accountable for results actually achieved" (17). Embedded in a web of competing legitimate expectations, PBC represents a switch away from a hierarchical accountability with input and process orientations toward a professional accountability that allows for the exercise of professional discretion and expertise to achieve targeted results (Romzek, 2000).

Generally, PBC is expected to promote better service outcomes. PBC, by making contract compensations attached to performance achievements, draws contractors' attention toward the results of service delivery, away from service delivery per se. The discussion on performance management is always based on the notion of "what gets measured gets done" – when people are given clearly measured targets, they would pay sufficient attention to achieve them. Following this line of reasoning, PBC would encourage service performance improvement. Further, as contractors are given much freedom in service process to prescribe services, the amount of administrative reporting and paperwork required by public agencies is greatly reduced. As such, contractors are believed to devote more time and energy to designing quality and innovative services to match client needs, which again enhances service outcomes. Combining these two together, PBC promises greater government acquisition efficiency, i.e., doing more with less. Under PBC, only contractor efforts that result in desired outcomes would be reimbursed, which maximizes the productivity of administrative resources. Less government monitoring also reduces administrative costs substantially.

In its essence, PBC stands for a marriage of service contracting with performance management, two prevalent managerial tools in contemporary public administrative narrative. On one hand, as mentioned earlier, service contracting has been a common and desired practice at virtually all levels of governments. Today, governments heavily collaborate with third-party nongovernmental actors to deliver various services through publicly funded contracts and grants. However, along with the

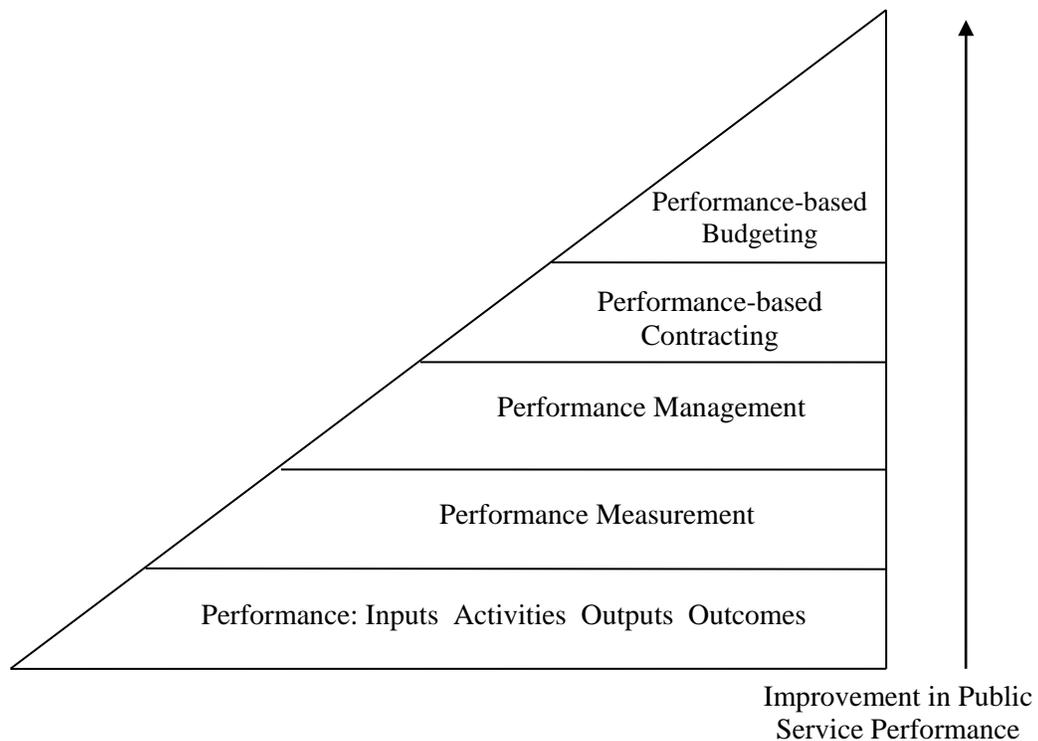
widespread use of contracting, contracting management is often found to be problematic. In this vein, PBC, by introducing performance measures into contracting management, can be seen as an endeavor in helping address this challenge.

On the other hand, PBC is an extension of government performance management strategy. Although performance measurement in government management appeared as early as the beginning of the twentieth century (Williams, 2003), the popularity of “performance” in public administration discourse is largely due to the Government Reinventing movement in the early 1990s (Kettl, 2005; Radin, 2006). The Government Performance and Results Act (GPRA), drawing government attentions on federal programs away from rules and process to results, service quality, and customer satisfaction, became the prelude of the nationwide performance movement. Gradually, governments at all levels started to adopt performance measures in resource allocation and program management and establish a variety of pay-for-performance systems to align budgetary and managerial decisions with performance achievements (e.g., Behn, 2003; Hatry, 2006; Heinrich, 2002; Joyce, 1993; Kravchuk & Schack, 1996).

At the outset, performance management activities were mostly run within government organization domain. However, as public administration evolves, more and more indirect government tools (e.g., contracts, grants) are introduced into the governance system (Salamon, 2002). Public administration today is no longer a tale of government, but more of governance (Kettl, 2002). This implies that government

performance depends on not only direct government tools, but indirect ones. As Frederickson and Frederickson (2006) show, the strength of an agency performance nowadays is deeply embedded in the characteristics of third-party grantees and contractors. With the nationwide performance movement keeping reshaping and redefining the structure and process of public administration activities, it is inevitable to witness the expansion of performance elements to the management of indirect government tools, forming a relatively comprehensive government performance management system. PBC thus becomes an indispensable part therein (See Figure 2).

Figure 2. Performance Reform Hierarchy



Source: Smith & Grinker (2004).

As said, the extensive use of government contracting as an indirect government tool to deliver products and achieve policy goals has fundamentally redefined the U.S. governance system, in both political and managerial senses. In this way, the management of contracting process to ensure high-quality results becomes an imperative challenge. Unfortunately, public management literature has documented that public managers at all levels of governments fail to address this challenge effectively. As a response, public management scholarship and practice in recent decades have conducted a huge amount of exploration of effective contracting management strategies. Inspired by performance management movement, PBC represents one of the most recent efforts. By attaching contract compensations to contract performance, rather than the delivery of service per se, PBC promises better outcomes, less service costs and administrative monitoring. Given these potential benefits, PBC is currently very popular in a variety of service areas and advocated by different levels of governments. However, PBC is not a brand-new managerial tool; its historical root could be dated back to two decades ago. Moreover, even with the historical evolvement of PBC in mind, the documented evidence on PBC effectiveness is still unclear. These are the topics of the next chapter.

## Chapter 2. Burgeoning Popularity, Different Models, and Elusive Effectiveness

### 2.1 The Evolution of PBC at Federal Level

Federal agencies have used PBC to varying degrees for acquiring a wide range of goods and services. Although PBC has been referred to in government regulations, guidances, and policies for about two decades, the historical root of PBC in the federal government can be dated back to even earlier. For example, in the early 1970s, the Office of Economic Opportunity (OEO) in the Department of Health, Education, and Welfare attempted to introduce PBC in educational services. Some school districts contracted out some portion of their instructional activities with private companies and attached contract payment to the extent to which contractors helped students learn (Gramlich & Koshel, 1975; Mecklenburger, 1972; Levine, American Educational Research, & American Association of School, 1972). The results of the initiative were quite mixed and problems arose in the implementation process. Participating organizations failed to reach consensus on several important questions such as the validity of standardized tests as achievement measures and what should be measured. The efforts of introducing PBC to educational services in this experiment were soon dropped.

Despite the early trial, federal implementation of PBC was not fully pursued until the Congress and the Office of Management and Budget (OMB) expressed enough

enthusiasm. Overall, the exploration of PBC in the federal government formally began in 1990s, represented by the appearance of the Office of Federal Procurement Policy's (OFPP) Policy Letter 91-2 on Service Contracting. The policy letter believed that PBC "enhances the Government's ability to acquire services of the requisite quality and to ensure adequate contractor performance," and advocated that all federal agencies should "use performance based contracting methods to the maximum extent practicable when acquiring services". In 1994, OMB initiated a governmentwide pilot project to encourage the use of PBC in federal agencies. In 1997, the Federal Acquisition Circular 97-01 amended the FAR to implement OFPP policy letter 91-2 and confirmed the policy that PBC should be used as the preferred service acquisition method (FAR 37.102). The FAR currently establishes a policy that federal agencies use PBC to the maximum extent practicable for service acquisition.

This preference on PBC remains in recent years. In fiscal year 2001, federal agencies reported a \$28.6 billion use of PBC, 21% of the total obligations (\$135.8 billion) incurred for services (GAO, 2002). The Services Acquisition Reform Act of 2003 also lends its strong support for PBC. In fiscal years 2005-2007, federal agencies were required to apply PBC to 40% of eligible service actions, including contracts, task orders, modifications, and options. In fiscal year 2008, they were encouraged to expand their PBC efforts on eligible service actions to 50%. OFPP also mandated that federal agencies to submit performance-based acquisition agency-wide management plans for fiscal years 2007-2011, outlining their progress and plans in applying PBC

to eligible service contracts. Table 1 provides a brief roadmap of the historical evolution of PBC use in the federal government.

Table 1. A Brief History of PBC in the Federal Government

Year	Federal Agency/Act	Document
1980	OFPP	A Guide for Writing and Administering Performance Statements of Work for Service Contracts
1991	OFPP	Policy Letter 91-2
1993		Government Performance and Results Act
1994	OFPP	Performance-Based Service Contracting Pledge
1997	OFPP	Memo on “Performance-Based Service Contracting Checklist”
1997	FAC (Federal Acquisition Circular)	FAC 97-01
1998	OFPP	Report on Performance-Based Service Contracting Pilot Project
1998	OFPP	Best Practices for Performance-Based Service Contracting
2000	National Defense Authorization Act FY 2001	Statutory Preference for Performance-Based Service Contracting
2001	FAC	FAC 97-25
2001	OMB	Memo on “Performance Goals and Management Initiatives for FY 2002 Budget”
2002	FAC	FAC 2001-07
2002	GAO	Report on “Guidance Needed for Using Performance Based Service Contracting”
2003	OFPP	Report on “Performance-Based Service Acquisition: Contracting for the Future”
2004	OFPP	Memo on “Increasing the Use of Performance-Based Service Acquisition”
2006	OFPP	Memo on “Use of Performance-Based Acquisitions”
2007	OFPP	Memo on “Using Performance-Based Acquisition to Meet Program Needs – Performance Goals, Guidance,

		and Training”
2007	OFPP	Memo on “Fiscal Year 2008 Performance-Based Acquisition Performance Goal”

## 2.2 Popularity at State and Local Levels

State and local governments have also shown growing interest in using PBC in the purchase of goods and services. Almost every state has introduced PBC in their acquisition efforts to some extent. Although there is no uniform effort in state and local governments that responds to the federal initiatives, their explorations of PBC are much more dynamic and diverse. For example, Washington State issued Executive Order 10-07 on Performance-based Contracting in 2010, advocating the use of PBC. It requires all state agencies shall (1) require that new contracts for products and services meet performance-based contracting standard, (2) review existing contract prior to renewal and update as necessary to reflect performance-based contracting standards, and (3) ensure performance-based contracts are actively managed to meet performance-based standards.

Particularly in human services, the interest in PBC is expanding rapidly. In Maine, State Statutes mandate the use of PBC in all human service contracting (Me. Rev. Stat. Ann. tit. 22, § 214). In California, eight of the nine counties in Southern California are using PBC in services such as employment training, aging and adult, and juvenile services, with most tying contract payment to a set of defined service outcome milestones (Daly, Tucker-Tatlow, & Gibson, 2004). New York City is demonstrating a growing commitment to PBC in its human service contracts. Most

human service contracts there have already included performance indicators and linked contract payment or renewal to contractor achievement in these indicators (Krauskopf, 2008). Although the detailed PBC designs in these states may vary, the motivations behind the injection of PBC into state efforts in service acquisitions are basically the same: to help align human service systems' focus on outcomes with how services are financed. Through restructuring contract specifications and compensations, human service agencies bind contractors with their service outcomes and maximize service acquisition efficiency. Overall, PBC is widely used in four human service areas: substance abuse treatment, child welfare, mental health, and employment training. The following discussion provides some documented evidence of PBC use in these fields. Although the survey here could not be exhaustive, it does make sense of the current status of PBC in human service provision.

### ***Substance Abuse Treatment***

The U.S. Institute of Medicine has advocated the use of performance measures in payment systems to promote quality improvement in treatment services since early 1990s (Institute of Medicine, 1990). The institute reiterates this suggestion in a number of its later reports (Institute of Medicine, 2001; 2006). So far, at least two states have formally responded to this call and their practices have been well documented.

Maine was the first state to include PBC in its purchase of addiction treatment services. In 1992, the Maine Office of Substance Abuse launched a PBC system to finance all publicly funded substance abuse treatment services (Commons et al., 1997). Under the PBC system, all programs were evaluated on post-treatment patient indicators within three categories: effectiveness measure (the minimum percentage of discharged clients who had achieved certain outcomes, such as abstinence and employment), efficiency measures (the units of treatment that providers had to deliver, such as number of clients served, number of services per client), and special populations (the targeted percentage of difficult clients, such as homeless people and youths). Each contract within the system would specify a minimum standard on each indicator that a contractor has to satisfy. Those contractors who failed to meet the minimum expectations might incur corrective actions and financial penalties.

Table 2 PBC in Maine Substance Abuse Treatment Services

	<i>Outpatient</i>	<i>Residential Rehabilitation</i>	<i>Detoxification</i>
<i>Efficiency Standards</i>			
Minimum service delivery (percent of contracted amount)	90%	80%	70%
Minimum service delivery to primary clients (percent of total units delivered)	70%	N.A.	N.A.
Number to be met	2 of 2	1 of 1	1 of 1
<i>Effectiveness Standards</i>			
Abstinence/drug free 30 days prior to termination	70%	85%	N.A.
Reduction of use of primary substance abuse problem	60%	85%	N.A.
Maintaining employment	90%	90%	N.A.

Employment improvement	30%	5%	N.A.
Employability	3%	3%	N.A.
Reduction in number of problems with employer	70%	N.A.	N.A.
Reduction in absenteeism	50%	N.A.	N.A.
Not arrested for OUI offense during treatment	70%	N.A.	N.A.
Not arrested for any offense	95%	N.A.	N.A.
Participation in self-help during treatment	40%	80%	N.A.
Reduction of problems with spouse/significant other	65%	60%	N.A.
Reduction of problems with family members	65%	60%	N.A.
Referral in continuum of care	N.A.	90%	45%
Referral to self-help	N.A.	N.A.	20%
Time in treatment	N.A.	N.A.	4 days
Number to be met	8 of 12	5 of 9	2 of 3
<i>Special Populations Standards</i>			
Females	30%	40%	14%
Age: 0-19	10%	4%	1%
Age: 50+	6%	5%	12%
Corrections	25%	10%	2%
Homeless	1%	1%	20%
Concurrent psychological problems	8%	3%	11%
History of IV drug use	12%	15%	27%
Poly-drug use	35%	40%	28%
Number to be met	5 of 8	5 of 8	5 of 8

Note:

1. Percentages are the minimum percent of total clients that must meet the indicator for the program to be deemed to have met that indicator.
2. N.A. means that programs offering the treatment modality are not required to meet the indicator
3. Number to be met is the number of indicators the program must meet to be deemed to have performed in that category.

Source: Commons et al. (1997).

More recently, the Delaware Division of Substance Abuse and Mental Health changed its contracting method in alcohol and other drug treatment programs from a FFS basis to a PBC basis in 2001. Under PBC, contractors were paid monthly based on their performance on three performance measures – utilization of treatment capacity, client participation in treatment, and client treatment completion (McLellan, Kemp, Brooks, & Carise, 2008; Stewart, Horgan, Garnick, Ritter, & McLellan, 2013). Table 3 shows the performance measures and payment schedule for the first two indicators. In addition, providers, after helping clients complete treatment (i.e., active participation in treatment for a minimum 60 days, achievement of treatment goals, and a minimum 4 consecutive weeks free from alcohol and illegal drugs) may receive \$100 bonus per client.

Table 3 PBC in Delaware Alcohol and Other Drug Treatment Programs

Program Capacity Utilization			Treatment Participation Requirements			
Target rate 2001-2002	Target rate 2003-2007	Payment: % of contract amount	Client treatment phase	Client treatment participation requirement	% clients required to meet target	Payment: % of contract amount
80%	90%	100	Phase 1	2 visits/week	50	1
70%-79%	80%-89%	90	Phase 2	4 visits/month	60	1
60%-69%	70%-79%	70	Phase 3	4 visits/month	70	1
50%-59%	60%-69%	50	Phase 4	2 visits/month	80	1

Note: treatment participation payments are conditional on achieving the capacity utilization requirement. Additional 1% payment when the program meets all four participation target.

Source: Stewart et al. (2013).

### *Child Welfare*

Child welfare might be the area where PBC enjoys the most attention and praise. The traditional fee-for-child contracting was found to undermine permanency: once a child welfare issue has been resolved and a child has been discharged, a contractor would face revenue loss unless a new child is referred. Thus, contractors may be inclined to keeping child in care rather than moving them toward permanency. Since 1990s, child welfare agencies have experimented PBC to purchase a variety of services, such as adoption, foster care case management, in-home services, residential care, and so on. In fiscal year 2008-2009, 24 states reported that their lead agencies include in service contracts benchmarks or indicators to measure service accessibility, timeliness, and service delivery efficiency (U.S. Child Care Bureau, 2008). Within the same time period, the Quality Improvement Center on the Privatization of Child Welfare Services found 14 states had service contracts that directly connect contract payment to performance and 11 states would consider contractor's performance achievement when making future funding decisions. In 2005, the Children's Bureau at the Department of Health and Human Services funded a project to test the use of PBC in child welfare services in Florida, Illinois, and Missouri.

Table 4 State Use of PBCs in 2009

	Operational Definition	States	Number
PBCs link contractor payment to performance	States with at least one PBC that links payment to performance, most commonly in the way of	AZ, FL, IA, ID, IL, MI, MN, MO, NC, ND, NE,	14

	service or client outcomes	NM, TN, WY	
PBCs inform contract renewal decisions	States using performance measures in contracts primarily to gauge contract renewal decisions	AK, AR, CA, CO, CT, IN, LA, OH, OR, WA, WI	11

Source: The Quality Improvement Center on the Privatization of Child Welfare Services, 2009.

However, there is a significant variance in the detailed PBC design across states, in terms of performance measures, payment structures, and other dimensions. Table 5 provides a snapshot of the current state of PBC use in some states.

Table 5 PBC in Selected State Child Welfare Agencies

State	Contracted services	Geographic coverage	PBC initiated	Selected performance measures
FL	Foster care	Judicial circuit 5	2007	<ul style="list-style-type: none"> <li>• Earlier and more accurate data entry into state's administrative system</li> <li>• Increased contracts with biological parents</li> <li>• Improved rates of maintained permanency of children</li> </ul>
IA	Resource family recruitment	statewide	2007	<ul style="list-style-type: none"> <li>• Sufficient pool of foster and adoptive homes</li> <li>• Children matched with appropriate foster homes in a timely manner</li> <li>• Safety in foster and adoption care</li> </ul>
IL	Foster care case management	statewide	1998	<ul style="list-style-type: none"> <li>• Child safety (e.g., #of reports of abuse/neglect)</li> <li>• Child well-being (e.g.,</li> </ul>

				<ul style="list-style-type: none"> <li>placement of siblings, placement within community</li> <li>• Child permanency (e.g., average length of stay in care, placement disruption)</li> </ul>
IL	residential care and treatment	statewide	2008	<ul style="list-style-type: none"> <li>• Sustained favorable discharge rate</li> <li>• Treatment opportunity days rate</li> </ul>
IL	independent living and transitional living programs	statewide	2009	<ul style="list-style-type: none"> <li>• Discharge potential rate with indicators of self-sufficiency</li> <li>• Transitional living placement stability rate</li> </ul>
MO	foster care and adoption case management	Three regions	2005	<ul style="list-style-type: none"> <li>• Reduced reentry into foster care</li> <li>• Increased stability</li> <li>• Increased permanency</li> </ul>
NM	Adoptive and foster home licensing	statewide	2008	<ul style="list-style-type: none"> <li>• Home studies completed in a timely manner</li> </ul>
TN	Foster care case management	statewide	2007	<ul style="list-style-type: none"> <li>• Average care days</li> <li>• Proportion of placements existing to permanency</li> </ul>
WY	Residential treatment	statewide	2006	<ul style="list-style-type: none"> <li>• Reduced length of stay</li> </ul>

Source: Child Care Bureau. (2009).

### ***Mental Health***

Mental health is one of the pioneers in human service areas that experimented PBC.

As early as late 1970s, state mental health agencies have tentatively introduced performance measures into their service acquisition efforts. Wisconsin was the first state to initiate PBC with localities for mental health care in 1973. In Wisconsin, each

local mental health authorities (LMHA) received a fixed budget from the state for community treatment and for state hospital treatment. LMHA was responsible for all costs incurred in the provision of services to its population. Community care costs were borne directly by LMHA, either through its own provision of services, or through the costs of contracts for the provision of services. LMHA was charged for state hospital use at per unit cost. Since LMHA received a fixed amount from the state, it received a bonus if their usage fell below this target, and a penalty for usage above the target (Chapin & Fetter, 2002; Gaynor, 1990). Michigan also adopted this model later.

In fiscal years 1978-1979, the Division of Mental Health in Colorado introduced PBC into its mental health system (Glover & Berger, 1989; Miller & Wilson, 1981). When contracting with community mental health centers for mental health services, the state agency included several categories of performance indicators (such as number of admissions by age group, regular reporting of the pre- and post- outcome on all clients, number of severely disabled to be served, contractor's accomplishment in Affirmative Action Plan) in their service contracts. At the end of the contract year, contractors had to report their achievement in these categories. A failure to serve 93% of the categorical quotas might result in a 5-7% reduce in contract funding for the next year.

The Philadelphia mental health residential system started PBC experiment in late 1990s, aiming to elevate low occupancy rates and prioritize access to residential care

for persons with the greatest needs. Before that, service contractors were compensated based on the availability of residential beds. This contracting method was found to discourage the efficient use of resources and lead to chronically low occupancy levels. In 1998, the Occupancy Based Reimbursement system was launched, directly tying occupancy performance to financial incentives and sanctions. Service contractors were required to maintain annualized occupancy rate of 86% at a minimum to avoid financial sanctions (not exceeding the equivalent of 3% of the program's yearly costs). Programs that maintained annualized occupancy levels of 93% or higher could receive incentive funds (not exceeding the equivalent of 3% of the program's yearly costs). Starting 2004, client outcome measures (e.g., graduation and hospitalization rates) were introduced into the PBC system (Faith et al., 2010).

### ***Employment Training***

Employment training programs also have a very long history of using PBC. For example, employment programs funded by the Job Training Partnership Act included client-level performance measures in their contracts and made funding decisions based on performance achievements (Barnow, 2000; Heckman, Heinrich, & Smith, 2003). The Workforce Investment Act (WIA), JTPA's successor as the primary federal training program, adopted an expanded version of the JTPA performance system. Besides, Wisconsin transferred its Wisconsin Works (W-2) contracts from a cost-reimbursement basis to a PBC basis in 1997, tying contract payment to measured performance. In this PBC system, detailed performance measures changed over time.

Contractors who failed to meet basic performance standards might lose future contracts, while capable contractors would enjoy profits or bonuses (Heinrich & Choi, 2007) .

Particularly, employment services for disabled people within state vocational rehabilitation programs are increasingly using PBC in the purchase of various services from contractors. Since Oklahoma designed and used the milestone payment system (one version of PBC) in 1990s, many other states such as Alabama, Indiana, Massachusetts, and New York have followed the lead (O'Brien & Revell, 2005). The details of state vocational rehabilitation programs and their PBC models in employment service contracting will be presented in depth in chapter four.

### 2.3 Important but Missing Links

Despite the burgeoning popularity in the use of PBC to purchase human services, there is not much documented evidence on the effectiveness of PBC. Specifically, two critical issues related to the use of PBC in human services remain unclear: (1) whether PBC produces better results than fee-for-service contracting – the effectiveness problem, and (2) if so, under what conditions, or how to use or implement PBC – the capacity problem.

#### ***Effectiveness Problem***

Ironically, to date, there is still little empirical evidence supporting that PBC actually leads to performance improvement in human services in a systematic way. The current prevalence of PBC in the purchase of human services is largely driven by the underlying theoretical reasoning behind PBC and the fashion of PBC in other fields. The theoretical reasoning of PBC is tempting: attaching contract compensations to service outcome measures could motivate better outcomes and empowering contractors could encourage innovative and quality services. The effectiveness of PBC in other fields such as energy further makes PBC attractive to human service agencies. For example, the federal government conducted a performance-based service contracting pilot project in 1998 and found a 15% decrease in contract prices and a 18% improvement in customer satisfaction (OFPP, 2003). However, current fad of PBC in human services often ignores the distinct characteristics human services possess and the special challenges those features bring to PBC. The discussion on this point is relatively brief here; a more detailed theoretical elaboration will be found in the next chapter.

The foremost precondition of PBC is the inclusion of performance measures. Any performance-based management tool requires a set of performance standards and metrics against which success could be measured. However, developing comprehensive and quantifiable measures that could cover the full spectrum of human service performance has long been considered very tough, if possible. First, human service programs frequently pursue values or goals that are multi-dimensional and often competing, which makes the design of appropriate measures that could perfectly cover the full range of the missions and values very difficult (Behn, 2003;

Heinrich & Fournier, 2004). Second, human service outcomes cannot easily be attributed to particular interventions and the confounding factors would contribute to the ambiguity of outcomes. Third, most human service programs aim to promote long-term stability and positive quality-of-life changes, but performance measures in service contracts have to emphasize short-term effects within certain contract duration. As a result, public managers have to use intermediate outcomes to account for final outcomes (Martin & Kettner, 1996). In short, all these elements jointly imply that performance measures for human services are often biased.

In addition to the problem of ambiguous performance, human services also feature high provider discretion in service delivery process. Human service provision is highly labor intensive, making the exercise of discretionary judgments by service providers inevitable or even desired (Lipsky, 1980; Riccucci, 2005; Sandfort, 2000). The line staff, through direct interactions with clients, can determine the “range of behavioral actions from which clients may choose their responses” (Lipsky, 1980, 61). Thus, such discretion constitutes part of service providers’ daily work, actually playing a double-edged role. On one side, it can help providers “process” clients in a responsive way, tailoring services to different client situations. On the other hand, there is a risk that such discretion might be abused without justification.

In sum, the rise of PBC in human services represents the convergence of imperfect performance measures and high provider discretion. Combining these two together, human services indeed bring challenges to PBC and make it at the risk of “rewarding

A, while hoping for B” (Kerr, 1975). Relying on imperfect surrogate measures leaves service contractors room to “gaming,” while higher provider discretion granted by PBC helps contractors achieve these potential gains (Bevan & Hood, 2006; Bohte & Meier, 2000; Heckman, Heinrich, & Smith, 1997; Moynihan, 2011). For example, in several human service areas, when contract payments are tied to clients’ outcome achievement, contractors are likely to select clients and serve those who are easier to meet performance goals. Thus, PBC creates much potential for service contractors to “gaming” or “creaming,” by focusing services on the variable measured, while excluding other outcomes which may be equally important but more difficult to measure. As Radin (2006) suggests, “because various players are likely to use the information to meet their varied agendas, it is rational for those who are the subject of the data to find ways to game the system” (207-208).

Actually, current evidence on the effectiveness of PBC in human services, though limited and unsystematic, has already been quite mixed. Table 6 demonstrates some of these studies. The introduction of PBC into substance abuse treatment programs has attracted strong scholarly interest in examining various aspects of its effectiveness. Commons, McGuire, and Riordan (1997) compare the client-level changes before and after the use of PBC in Maine and observe positive improvement in service outcomes, such as abstinence, reduction in drug use, reduction in problems with jobs, and no arrests. However, this finding was largely doubted by later studies in that it fails to consider the unintended effects incurred by PBC. Shen (2003) finds that, after the implementation of PBC in Maine addiction treatment, the number of most severe

clients dropped by 7% and concludes that PBC actually equips contractors with financial incentives to treat less severe clients to achieve targeted performance. Lu (1999) argues that since state agency relied on contractors to report client treatment outcomes, contractors had incentives to misreport and cheat on performance information to ensure funding from state government. Brucker and Stewart (2011) reexamine Maine's experience and conclude that PBC had no positive effect on program performance such as time to treatment, level of client participation, length of stay, and completion of treatment. In Delaware, McLellan et al. (2008) find significant increases in average capacity utilization (from 54% to 95%) and average proportion of patients' meeting participation requirement (from 53% to 70%) after PBC implementation, with no notable demographic changes in the patient population over time. Building on this finding, Stewart et al. (2013) further trace the effectiveness of PBC on individual clients and observe 13 days less in waiting time for treatment and 22 days longer in length of stay in treatment.

In employment services, the effectiveness of PBC in the programs funded by the Job Training Partnership Act has been found to be very controversial. The use of short-term and straightforward measures is only weakly, and sometimes perversely, associated with long-term welfare (Barnow, 2000; Heckman et al., 2003; Heinrich, 1999). Dias and Maynard-Moody (2006) study workers in a for-profit subsidiary of a national marketing research firm that shifted into the business of providing welfare services. They find requirements for meeting contract performance (job placement) and profit quotas created considerable tensions between managers and workers on the

importance of meeting performance goals versus meeting client needs. The easiest way to meet contract goals and gain profits was to minimize the time and effort devoted to each client. Koning and Heinrich (2013) examine the incentive effects of PBC on program outcomes in Dutch welfare-to-work program. They find evidence of gaming activities, but these activities had little impact of gaming on service outcomes. They conclude that the use of PBC increased job placement, but not job duration.

In other human service areas, the effectiveness puzzle remains. Many evaluations of PBC in child welfare are still underway. In particular, Illinois used PBC to promote permanency outcomes in its foster care contracting and witnessed a significant decrease in the number of children in out-of-home placement (Kearney, McEwen, Bloom-Ellis, & Jordan, 2010). After Philadelphia directly tied financial incentives and sanctions to occupancy performance, the mental health residential system witnessed a significant increase in occupancy, with an average occupancy rate of mid 90%. However, there was still a concern that the performance target on occupancy may suppress the flow of residents through the housing system (Faith et al., 2010).

Table 6. Selected Studies on PBC Effectiveness in Human Services

<b>Author(s)</b>	<b>Study site</b>	<b>Contracted services</b>	<b>Unit of analysis</b>	<b>Findings</b>
Commons, McGuire, and Riordan (1997)	Maine	Substance abuse treatment services	Client	<ul style="list-style-type: none"> <li>Improvement in service outcomes, such as abstinence, reduction in drug use, reduction in problems with jobs, and no arrests</li> </ul>
Lu (1999)	Maine	Substance abuse treatment services	Client	<ul style="list-style-type: none"> <li>Providers had incentives to report better treatment performance</li> </ul>

				outputs
Heinrich (1999)	Chicago	JTPA programs	Program	<ul style="list-style-type: none"> <li>• Performance measures were not strongly correlated with program goals</li> <li>• Cost-per-placement measure had negative implications for service quality</li> </ul>
Shen (2003)	Maine	Substance abuse treatment services	Client	<ul style="list-style-type: none"> <li>• Number of most severe clients dropped by 7%</li> </ul>
Lu, Albert Ma, and Yuan (2003)	Maine	Substance abuse treatment services	Client	<ul style="list-style-type: none"> <li>• More referrals and better match between illness severity and treatment intensity</li> <li>• A positive but insignificant effect on dumping (a client is sequentially referred from one provider to the next without being treated)</li> </ul>
Heckman, Heinrich, and Smith (2003)	US nationwide	JTPA programs	Client	<ul style="list-style-type: none"> <li>• Short-term measures were weakly, even perversely, related to long-term impacts</li> <li>• Efficiency gains or losses from gaming were small</li> </ul>
Dias and Maynard-Moody (2006)	Porter City	Welfare-to work program	Program and client	<ul style="list-style-type: none"> <li>• Distorted incentive structures that led to programmatic conflicts between program management and staff</li> <li>• Negative program practice and poor client outcome</li> </ul>
Heinrich and Choi (2007)	Wisconsin	Wisconsin Works (W-2) program	Program	<ul style="list-style-type: none"> <li>• Contractors responded to performance incentives related to future funding decisions</li> <li>• Insufficient contacting management may undermine PBC effectiveness</li> </ul>
McLellan, Kemp, Brooks, and Carise (2008)	Delaware	Outpatient alcohol and other drug treatment	Program	<ul style="list-style-type: none"> <li>• Average capacity utilization rates increased from 54% to 95%</li> <li>• Average proportion of patients' meeting participation requirement</li> </ul>

				increased from 53% to 70%
Faith et al. (2010)	Philadelphia	mental health residential services	Program	<ul style="list-style-type: none"> <li>• Significant increases in program occupancy</li> <li>• The flow of residents through the housing system might be suppressed</li> </ul>
Stewart, Horgan, Garnick, Ritter, and McLellan (2013)	Delaware	Outpatient alcohol and other drug treatment	Client	<ul style="list-style-type: none"> <li>• Waiting time for treatment declined 13 days</li> <li>• Length of stay in treatment increased 22 days</li> </ul>
Koning and Heinrich (2013)	Netherlands	Welfare-to-work services	Client	<ul style="list-style-type: none"> <li>• Evidence of gaming activities</li> <li>• Little impact of gaming on service outcomes</li> </ul>

Overall, current research on the effectiveness of PBC in human services mostly suffers from two limitations. First, many studies fail to count in the impact of unintended consequences of PBC on full service performance. As mentioned above, developing a series of performance measures that could capture full service performance is very challenging. Thus a common strategy is to use short-term and easy-to-measure indicators instead. As such, contractor efforts in achieving measured performance may affect their behaviors related to unmeasured performance. For example, the performance improvement in Maine substance abuse treatment programs was very likely to be attained though custom selection and contractor misreporting. Such performance improvement, though efficient to some extent, should not be considered effective. More broadly, if improvement in measured performance is achieved at the expense of other unmeasured performance, such improvement is not effective and desired. A systematic evaluation of PBC effectiveness should include such consideration. Without it, the evaluation is inevitably biased.

Second, methodologically, these evaluation studies often rely on “pre-post” comparisons based on observation data. The most severe threat to internal validity in observation studies is that observations in comparison groups are biased by counterfactual variables, which are not directly comparable. The “pre-post” comparison, as the most basic quasi-experimental design, is very unlikely to rule out the effect of these counterfactual variables (Shadish, Cook, & Campbell, 2002). Thus, the results from pre-post comparisons generally suffer from low internal validity. In this sense, more robust research designs should be used.

### ***Capacity Problem***

Closely related to the effectiveness of PBC is the capacity challenge. As discussed previously, PBC is experimented in and introduced to service contracting as an effort to address the smart-buyer problem, i.e., public managers are sometimes not equipped with sufficient management capacity to use contracting effectively. However, although the potential benefits of PBC are attractive, the launch of PBC system does not guarantee the achievement of those benefits. Rather, PBC itself creates a series of new challenges for public managers in designing and implementing PBC systems, such as how to set performance milestones and indicators, how to split responsibilities and risks between contracting parties, how to conduct performance monitoring, etc. After reviewing the use of PBC in federal agencies, GAO (2002) raises the concern that “whether agencies have a good understanding of performance-

based contracting and how to take full advantage of it” (2). New York State piloted PBC in its employment services for disabled people in early 2000s and soon abandoned the effort when the administration found they lacked the capacity to implement PBC and lead organizational change (Gates et al., 2004). Heinrich and Choi (2007) admit that insufficient program administration and contracting management capacity undermined the effectiveness of PBC in Wisconsin Works program.

Basically, the introduction of PBC requires two managerial capacities: designing appropriate PBC systems and implementing organizational changes. First, the critical role of performance measures could not be emphasized more. As is shown previously, there are many variations in performance measures among the PBC models currently used in different states, even in the same human service field. Appropriate performance measurement facilitates PBC implementation and reduces the potential of unintended consequences. This further implies several more detailed tasks, such as which part of performance to track and how to link contract reimbursement to client outcomes.

In addition to these technical aspects of PBC design, a more profound capacity would be leading organizational innovation and changes in an inter-organizational setting. Given the difficulty of designing comprehensive measurement systems for human services, this capacity becomes even more critical. Under PBC, only service efforts that successfully achieve desired outcomes would be reimbursed. Thus, PBC actually

forces contractors to burden substantial fiscal risks. Contractors are exposed to loss when their service efforts do not result in expected outcomes. Such risk shifting complicates contract implementation. Romzek and Johnston's (2002) study of service contracting in Kansas finds that although accurate performance measures in contracting may facilitate contract implementation, substantial risks at the contractor side would "compromise the capacity of the contractor both to meet performance expectations and to provide required performance information to contract managers" (430). McGrew et al. (2007) observe that contractors do prefer FFS over PBC, although they mostly welcome the freedom in the service process under PBC. In this sense, it is likely that contractors resist the transition from traditional FFS approach to PBC, or only perversely adjust to PBC systems.

Moreover, the injection of PBC to human service system is an evolving process, allowing longtime trial-and-error. The movement toward PBC takes patient and deliberate effort and needs to address a myriad of challenges. It is an evolutionary rather than a revolutionary process, which requires years' planning with progressive implementation and is expected to continue evolving over time. For example, over a 6-year period, the Philadelphia mental health system was able to shift from a FFS model to a PBC model. Even though the basic PBC framework had been there, the administration was still modifying and improving the performance measures (Faith et al., 2010). Particular, it takes a great deal of time to establish a meaningful performance measurement system that informs program development and client improvement. Public managers have to confront this evolutionary dynamic. As

Heinrich and Marschke (2010) argue, “an incentive designers’ understanding of the nature of a performance measure’s distortions and employees’ means for influencing performance is typically imperfect prior to implementation,” and thus “it is only as performance measures are tried, evaluated, modified, and/or discarded that agents’ responses become known” (203). All these imply that PBC should be treated as a learning process for public managers.

In sum, governments at all levels have shown substantial and continuous enthusiasm for PBC. In human services, particularly, state and local governments have expressed growing interests in using PBC in their service acquisition. Although the designs of detailed PBC systems in different states and different service areas might vary, the basic motivation of the injection of PBC is the same: align human service systems’ focus on outcomes with how services are financed, or more technically, reshape contractor behaviors through redefining contract incentive structures. However, the burgeoning popularity of PBC lacks sufficient evidence to show its promised benefits are actually achievable. The evidence available in this regard still fails to provide a consistent and persuasive answer.

To an extent, the introduction of PBC into human service systems, from a managerial perspective, needs to address the effectiveness problem (whether PBC produces better results) and the capacity problem (how to use PBC and lead interorganizational changes). The present research mainly focuses on the effectiveness problem, but would briefly discuss the implications on the capacity problem. Before that, the

research needs a theoretical framework that could pave the way for future discussion.

This is the topic of the chapter three: a theoretical discussion of contract design and its application to human service contracting.

## Chapter 3. Theoretical Framework

### 3.1 Formal Contract Design: A Principal-agent Perspective

The same as much previous literature on government contracting (e.g., Donahue, 1989; Johnston & Romzek, 1999; Kettl, 1993; Milward & Provan, 1998, 2000; Romzek & Johnston, 2005), this research puts the discussion of contract design first in a principal-agent model (Eisenhardt, 1989; Jensen & Meckling, 1976; Shapiro, 2005), where government (the principal) relies on contractors (the agents) to deliver human services and achieve policy goals. Based on the assumptions of goal conflicts and information asymmetry between the principal and the agent, the agency theory warns the existence of agency problem, i.e., the principal is subject to the agent's self-serving opportunistic behaviors. First, because of incomplete information, the principal could not verify the agent's capacity and thus may rely on low-quality agents. In this sense, it is the agent that chooses the principal, not the opposite. This is termed as adverse selection or hidden information (Arrow, 1984). Moreover, the agent may further take the information advantage to shirk his/her responsibility and not put forth the agreed-upon efforts. This hidden action (Arrow, 1984) would generate considerable moral hazard for the principal.

To address the agency problem, the principal might try a variety of monitoring tools to bridge information asymmetries and goal conflicts. However, all these efforts would incur agency costs. Therefore, the managerial implication of the agency theory

focuses on the design of efficient governance mechanisms to moderate the agency problem, or more precisely, appropriate control mechanisms to guide the distribution of risk and uncertainty between the principal and the agent. If organizational control is seen as a problem of information flow (Ouchi & Maguire, 1975), the design of control mechanisms and strategies within an organization largely rests upon two dimensions: (1) task programmability – the degree to which the means-ends relationships involved in agent behaviors can be precisely defined, and (2) outcome measurability – the extent to which various aspects of task outcomes could be specified in a comprehensive and quantifiable manner. The focus of control, therefore, can be on either the behavior of employees or the outcomes of those behaviors. Accordingly, the control strategy can be either behavior or outcome based (Eisenhardt, 1985; Ouchi, 1980; Thompson, 1967).

Generally, behavior-based control is appropriate in an environment characterized by high task programmability. When certainty regarding causation is high, control strategies are more reflected in high levels of monitoring and direction in agent activities, with performance evaluation often focusing on job inputs. If outcome measurability is high, organizations would prefer outcome-based control strategies, under which compensation schemes are attached to outcome measures and monitoring of employees becomes relative less. When a task is neither programmed nor measured, formal control mechanisms, both behavior-based and outcome based, seem ineffective in that there is no exact place to host the control. In this case, social control, or what Ouchi (1980) calls “clan” control, may emerge to play a

supplemental role. The social control system, using informal and normative mechanisms (such as shared values and norms of reciprocity) to align the preferences between the principal and the agent, implicitly encourages appropriate behaviors that could lead to desirable organizational outcomes.

Table 7. The Determinants of Organizational Control Strategies

		Task Programmability	
		High	Low
Outcome Measurability	High	Behavior or outcome control	Outcome control
	Low	Behavior control	“Clan” control

Source: Ouchi (1980).

Arrow (1964) defines the design of control strategies as the choice of operating rules and the choice of enforcement rules to support the operating rules. If an organization operates “as a nexus for a set of contracting relationships among individuals” (Jensen and Meckling, 1976, 310), then the design of optimal contract arrangement governing the principal-agent relationship constitutes the enforcement rule to facilitate contract implementation. In accordance with two types of organizational controls, there are two major contract alternatives: behavior-based and outcome-based contracts. The choice of a contract type is thus a function of task programmability and outcome measurability. The key in structuring contractual relationships, writes Eisenhardt (1989), is “the trade-off between (a) the cost of measuring behavior and (b) the cost of measuring outcomes and transferring risk to the agent” (61).

Figure 3 describes four types of goods and services in terms of their certainty in causation and outcome and different contract types tailored to fit these characteristics.

Figure 3. The Determinants of Contract Type

		Task Programmability	
		Low	High
Outcome Measurability	Low	Cell 1	Cell 2 Behavior-based Contracts
	High	Cell 3 Outcome-based Contracts	Cell 4 Behavior-based or Outcome-based Contracts

For services in Cell 2, the means-ends relationships involved in agent services can be explicitly specified and observed. As such, information asymmetry between the principal and the agent in terms of task programmability is low and the risk transferred from the agent to the principal becomes expensive. Therefore, the principal knows what the agent has done and could under behavior-oriented contracts to purchase the agent’s direct behaviors. In Cell 3, agent services are ambiguous to observe, but their outcomes could be clearly measured with less difficulty. Under these circumstances, the principal would prefer outcome-based contracts to align the agent’s incentives with those of the principal and make risk shifting from the agent to the principal become less likely. When both cause/effect relationships and outcomes

are highly certain (in Cell 4), there is no difference for the principal to control either service process or outcome, and thus both contract types work equally well.

The most problematic situation for contract design comes from the services in Cell 1, where agent services share both low task programmability and low outcome measurability. In health care, for example, the principal lacks the ability to anticipate clearly the treatment process and outcomes. As such, the locus of control for the principal seems obscure, leaving a high degree of incompleteness in contract specification. When the control the principal uses to govern the contractual relationships is incomplete, as incomplete contract theory (Hart, 1988; 1989) predicts, the agent would enjoy “residual rights of control” and be at the advantageous position in ex post bargaining and the division of ex post benefits. In most cases, the agent could perform discretionary judgments in the circumstances that were not specified in initial contracts. These behaviors are very likely to incur moral hazard. In short, the incompleteness in task programmability and outcome measurability would make contract design challenging. Unfortunately, this is where human services usually fit in.

### 3.2 Formal Contract Design for Human Services

Human services generally feature low task programmability and low outcome measurability. The effort on task programmability in human services is always disturbed by high provider discretion in the service delivery process. Human service provision is highly labor intensive, making the exercise of discretionary judgments by

service providers inevitable or even desired (Lipsky, 1980; Riccucci, 2005; Sandfort, 2000). Although there are various operating rules and service manuals throughout the service process, in real situations service providers are always required to apply their judgment and make decisions contingent on detailed contexts. These line staff, through direct interactions with clients, can determine the “range of behavioral actions from which clients may choose their responses” (Lipsky, 1980, 61). Thus, typically, service providers “do not do just what they want or just what they are told to want. They do what they can” (Brodin, 1997, 24). Thus, such discretion constitutes part of service providers’ daily work, actually playing a double-edged role. On one side, it can help providers “process” clients in a responsive way, tailoring services to different clients. On the other hand, it may abuse such rights without justification. Sandfort (2000) examines the potential influence of the new public management and traditional public administrative practices on front-line actions in two local welfare offices and two private contractors in Michigan. She finds that neither performance-based management nor traditional bureaucratic directives have an impact on front-line practices in either type of agency. Instead, the most powerful determinants of street-level behaviors rest upon the collective beliefs of front-line staff, such as norms, shared knowledge of the organizational members.

In addition, the outcome of human services is often too uncertain to be defined clearly. Measuring the performance of human service programs has long been considered demanding. First, from the normative perspective, like many other public programs, human service programs frequently pursue values or goals that are multi-

dimensional and often competing, such as efficiency, equity, and representativeness, derived from the various expectations on government cherished by citizens. Wilson (2000) details this multidimensional nature and the dilemma of balancing them. At the very basic level, public welfare programs are always involved in the efficiency-equity puzzle, recognized by Okun (1975) as “the big tradeoff.” Thus, the answers to the question of “what to measure” are always ambiguous and competing. As such, figuring out appropriate measures that could comprehensively cover the full range of the missions and values can be difficult (Behn, 2003; Heinrich & Fournier, 2004; Heckman, Heinrich, & Smith, 1997).

Second, technically, human services are directed to improving service recipients’ welfare through behavioral interventions. As Hasenfeld (1983) observes, human services aim to “protect, maintain, or enhance the personal well-being of individuals by defining, shaping, or altering their personal attributes” (1). However, beyond such interventions, there might be a number of uncontrollable factors out of service providers’ reach that would lower the certainty of desired outcomes (DeHoog & Salamon, 2002; Martin & Kettner, 1996; Wedel & Conston, 1988). Thus, outcomes cannot easily be attributed to a particular intervention. Also, the standards on significant changes in welfare conditions before and after services are sometime controversial.

Third, most human service programs aim to promote long-term stability and welfare, but performance measures have to emphasize short-term effects. Tracking persons

over time is a costly activity and does not produce short-term feedback on the success of the program. Most programs use outcomes of participants measured at the time they complete the program, or within a short period thereafter (Martin & Kettner, 1996). Both measures are short-term in nature, which creates another puzzle that these performance standards misdirect activities by focusing on the criteria that may be not related to long-term goals. Heckman, Heinrich, and Smith (2002) find that in the JTPA system, short-term measures used to monitor performance were only weakly, even perversely, related to long-term impacts. Putting these three points together, we could have some understanding on why performance measurement in human service programs is so difficult.

With these intrinsic characteristics of human services in mind, let's move onto the discussion of contract design. Traditionally, human service contracts run on a fee-for-service basis, a behavior-based contract, where government directly controls the service process (such as inputs standards and service methods employed) in order to ensure the delivery of promised services. When a client comes to a human service agency for services, agency staff would determine the eligibility and prescribe the amount of services needed. After that, the human service agency buys this amount of services from service contractors. For example, a human service agency may purchase individual counseling services for a domestic violence offender at the rate of \$75 per hour, or group counseling in outpatient substance abuse treatment at \$20 per 15 minute increment. After the services are delivered and paperwork is approved,

service contractors are reimbursed for that amount of services delivered, based on the unit of services (e.g., per hour or per 15 minutes).

However, the task programmability of a human service is always tentative – it is difficult to predict initially what services could exactly lead to desired results due to ambiguous jobs and uncertain future events. Thus, government effort on task programmability under FFS might be offset by the discretion contractors enjoy in the service delivery process because they work directly with clients and have (or pretend to have) more information on clients' service needs. And the negotiation nature of human service contracting may further justify the existence of discretion. As DeHoog (1990) observes, human service contracting generally follows a special negotiation or cooperation logic. Due to limited market competition, ambiguous performance, and costly contracting monitoring (DeHoog, 1984; Schlesinger, Dorwart, & Pulice, 1986; Van Slyke, 2003), human service contracting does not usually rely on the classical competitive bidding model. Rather, human service contracts are mostly specified through negotiations between government buyers and contractors. This would no doubt complicate government effort on task programmability.

Another byproduct of low programmability of human services is that the link between task and outcome becomes broken: due to failure of clear task specification, the detailed services prescribed by government do not necessarily lead to desired outcomes. In this sense, contract compensation, independent of service outcomes, only encourages service delivery, demonstrating a “triumph of process over results”

(Kettner & Martin, 1993, 62). Given this, contractors have no incentive to improve service performance. Further, better performance may even mean economical inefficiency for them (Wulczyn, 2005). For example, improving service quality increases contractor costs for advanced facilities and staff training, which would not be reimbursed by government. Again, better services reduce client demands for feature services.

The new PBC approach, holding an outcome orientation, draws contractors toward service results and leaves them considerable flexibility in serving clients.

Theoretically, PBC would encourage innovative services, better outcomes, and less monitoring. However, these benefits are subject to two assumptions—PBC is not vulnerable to (1) measurement problem and (2) gaming by contractors (Behn & Kant, 1999; Bevan & Hood, 2006). Without meeting these two requirements, the effectiveness of PBC cannot be guaranteed.

As mentioned above, the performance of human service programs is very challenging to track. In most cases, performance measures for human services are often just approximations of the targeted outcomes, i.e., short-term measures representing long-term effects and easy-to-measure goals representing ambiguous goals. As a result, the measurement assumption becomes problematic in human service contracting.

However, when surrogate measures are used, as mentioned earlier, service contracts become incomplete, leaving room for contractors to seek gaming and other strategic behaviors (Radin, 2006). Baker (2002; 1992) shows the efficiency of incentive

contracting depends on the extent to which the performance measures used are aligned with the principal's objective. When the principal's objective is not contractible, i.e., unclear or immeasurable, alternative measures have to be adopted as proxies. If so, the incentives associated with those performance measures are inaccurate and nonoptimal, leading contractors to engage in unintended activities even if contractors are risk neutral. The more distortion is in performance measures, the lower is the incentive for desired objectives.

Indeed, such distortion becomes even severer when gaming enters the picture. As noted above, service contractors embrace discretion when delivering services and PBC even enhances such discretion. Thus, it is very likely that contractors use their information advantage to conduct perverse adjustment to performance measures in order to appear to be behaving well (Hood, 2006; Courty & Marschke, 2004; Moynihan, 2008; Radin, 2006). Williamson (1985) terms this phenomenon as opportunism, a "self-interest seeking with guile" (47), which includes a wide range of behaviors, such as shirking, cheating, and withholding important information. Bevan and Hood (2006) summarize three forms of gaming problem under PBC context—ratchet effects (restricting current output to gain undemanding future performance target), threshold effects (downgrading the output of those performing better than the target to meet the target), and output distortions (achieving targeted performance measures at the expense of unmeasured performance). All would limit the effectiveness of PBC in human services.

Almost two decades ago, Cragg (1997) questioned why PBC was not prevalent in human service programs. After examining the practices of PBC in job training programs under JTPA, he concluded “unless performance standards are carefully designed, problems of moral hazard may preclude the widespread use of performance incentives in government programs” (147). Although performance measurement techniques have been improved greatly since Cragg’s study, the problem he observed may persist as long as ambiguous performance and high discretion associated with human services continue. By and large, neither behavior-oriented nor outcome-oriented contracts fit seamlessly with human services (Table 8). Both might incur certain amount of agency costs in structuring and monitoring contractual relationships, which would further undermine the effectiveness of formal contract design.

Table 8. Contract Type for Human Service Contracting

	Fee-for-service Contracting	Performance-based Contracting
Control Strategy	Behavior-based control	Outcome-based control
Implementation problem	Low task programmability	Low outcome measurability
Limitations	Triumph of process over results	Surrogate measures; Gaming behaviors

### 3.3 Informal Contract Design: A Relational Contracting Perspective

Another line of literature that would cast light on the discussion here is relational contracting. Interorganizational relationships (IORs) always embrace two dimensions, structural and relational, and thus propose two streams of governing mechanisms

(Faems, Janssens, Madhok, & Van Looy, 2008; Ring & Van de Ven, 1994). The structural perspective considers a formal structural arrangement and its role in structuring interorganizational behaviors and performance, in order to “create a predictable collaborative environment that mitigates exchange hazards and facilitates coordinated action” (Faems et al., 2008, 1054). The relational perspective emphasizes informal relationship building, trust cultivation, and trustworthy behaviors. Thus, it “promotes a more relational governance strategy in which partners rely on trust to address issues of safeguarding and coordination” (Faems et al., 2008, 1054). This dichotomy actually follows the conventional wisdom of the interaction between formal and informal behaviors in organizational management.

Following this line of reasoning, formal contracting centers on “detailed, binding legal agreements that specify the obligations and roles of both parties in the relationship” (Vandaele et al. 2007, 240). In this most visible part of a contract, the attention would be on the design of comprehensive contract clauses to bind future contingencies. In this sense, contracting basically means two elements: “(a) rational planning of the transaction with careful provision or as many future contingencies as can be foreseen, and (b) the existence or use of actual or potential legal sanctions to induce performance of the exchange or to compensate for non-performance” (Macaulay, 1963, 56). Thus, the designs of formal contracts aim to reduce uncertainties in contracting process and make contractor behaviors more predictable.

In contrast, relational contracting literature questions the gap between contract doctrine and the empirical operation of the contract system in the real world. Organizations engaged in contracting often do not need to conduct rational contract planning and negotiation when the transactions are run within a setting of continuing relationships. Potential disputes are compromised in the way of keeping the relationship continues. The contracting process, as Macaulay (1985) argues, is not “a neutral application of abstract rationality,” but “operates at the margins of major systems of private government through institutionalized social structures and less formal social fields” (477). Underlying this observation is the notion of relational contracting, a type of contracting that reflects “the relations among parties to the process of projecting exchange into the future” (Macneil, 1980, 4). It can be seen as a logical extension of the bounded rationality represented by formal contracting. This line of research (Macaulay, 1963; Macneil, 1977) highlights the role of relational sanction and social interaction in understanding the incentives under the fulfillment of contractual agreements. A detailed comparison between formal and relational contracting is listed below in Table 9.

Table 9. Comparisons between Formal and Relational Contracting

	Formal Contracting	Relational Contracting
Perspectives about relations with vendors	<ul style="list-style-type: none"> <li>• Anticipate short-term relationship</li> <li>• Low risk/low trust</li> <li>• No expectation for altruistic behavior</li> </ul>	<ul style="list-style-type: none"> <li>• Anticipate long-term relationship, seek out trustworthy partners</li> <li>• High risk/trust</li> <li>• Expect altruistic behavior in the interest of the whole</li> </ul>

Market assumptions	<ul style="list-style-type: none"> <li>• Many vendors available</li> </ul>	<ul style="list-style-type: none"> <li>• Few potential vendors</li> </ul>
Contract writing	<ul style="list-style-type: none"> <li>• Detailed specification of benefits, burdens, rules, and rights</li> <li>• Monitoring for compliance</li> <li>• Reliance on legal remedies</li> </ul>	<ul style="list-style-type: none"> <li>• Comparatively ambiguous contracts with anticipation of adapting to changing circumstances</li> <li>• Social norms serve as principal mechanisms of mediation or control</li> <li>• Aversion to third-party, legal remedies</li> </ul>
Management style	<ul style="list-style-type: none"> <li>• Sanctions imposed as written</li> <li>• Low levels of contacts and coordination</li> <li>• Compliance as a key concern</li> </ul>	<ul style="list-style-type: none"> <li>• Sanctions and remedies not imposed but rather negotiated and mediated</li> <li>• Flexibility, solidarity, information sharing</li> <li>• Maintenance of relationship as a primary concern</li> </ul>
Service characteristics	<ul style="list-style-type: none"> <li>• Easy to define service tasks</li> <li>• Easy to evaluate service quality and vendor performance</li> <li>• Tasks do not require special investment or customization and involve standardized service production processes</li> </ul>	<ul style="list-style-type: none"> <li>• Ambiguity in defining service tasks</li> <li>• Difficult to assess service quality and vendor performance</li> <li>• Vendors are required to make special investments to satisfy buyers' customized needs</li> </ul>

Source: Beinecke & DeFillippi (1999), Lamothe & Lamothe (2012), Sclar (2000), Williamson (1985).

This relational exchange perspective in contracting management has actually received growing attention in public management literature (e.g., Brown, Potoski, & Van Slyke, 2006; Sclar, 2000). For example, scholars in recent years have proposed to use

stewardship theory to explain public service contracting (e.g., Dicke, 2002; Lambright, 2009; Van Slyke, 2007). Stewardship theory emphasizes the cooperation and trust nature in principal- agent relationships. As Davis, Schoorman, & Donaldson (1997) suggest, it “defines situations in which managers are not motivated by individual goals, but rather are stewards whose motivates are aligned with the objectives of their principals” (21). Stewardship theory becomes more relevant to government-nonprofit contracting, in which nonprofits are always believed to be social-mission driven and have weaker incentives to take advantage of asymmetric information in market exchange. Such mission/value alignment with government would moderate goal conflicts between contracting parties and prevent nonprofit contractors’ opportunistic behaviors in maximizing their financial interest and market value.

Indeed, relational contracting has special implications for service contracting. First, human service contracting has been found to follow a negotiation model, rather than the competitive bidding model proposed by the privatization literature, due to limited market competition, ambiguous performance, and costly monitoring (DeHoog, 1991; Johnston & Romzek, 1999; Sclar, 2000; Van Slyke, 2003). In this sense, informal social exchanges between contracting parties would play a significant role. Romzek and Johnston (2002) find that in Kansas social service programs, ongoing “negotiation and collaboration among contracting partners” (423) is necessary for effective contract implementation. Brown and Potoski (2004) show that even in refuse collection, where service attributes are relatively easier to measure and market

competition is rich, public managers still engage in a variety of informal network activities (such as hosting informal meetings with contractors and attending professional conferences) to promote competition and reduce information asymmetries.

Second, as mentioned earlier, human service contracting is always troubled by low task programmability and low outcome measurability, which complicates the design of formal contract arrangement. As such, social control, or what Ouchi (1979) calls “clan” control, may emerge to function as a supplement. The existence of informal socialization process against organizational rationality in organizational operation has been long acknowledged since the Hawthorne Studies. The social control system, using informal and normative mechanisms (such as shared values and norm of reciprocity) to eliminate interest and goal incongruence between the principal and the agent, implicitly encourages appropriate behaviors that could lead to desirable collaborative outcomes. Put together, the arguments here call for the inclusion of relational aspects of contracting, in addition to formal contracting endeavor.

However, the interaction between formal and relational components of contracting is still under scholarly debate: whether formal and relational contracting could function as mutually competing or enhancing mechanisms. The mutual exclusion view considers a hostile relationship between formal and informal contracting. From this perspective, efforts on legal maneuvers in formal contracts as safeguards against potential breaches would be interpreted as a sign of distrust and hinder relationship-

building between organizations (Bernheim and Whinston, 1998; Ghoshal and Moran 1996; Lyons and Mehta, 1997). However, the complementary-role perspective challenges this view. Poppo and Zenger (2002) and Goo et al. (2009) find that clear contract specification reduces risk in cooperation, which would promote repeated exchanges and further result in mutual dependence and trust. And trust emerging from prior collaborations would substitute for more elaborate formal contract provisions (Gulati, 1995). Informal relationship and mutual understanding could mitigate ex post informal flow and coordination, reducing the need for clear specifications (Dore, 1983; Zollo et al. 2002). Sclar (2000) even argues with relational contracting, “the formal contract or agreement is less important as a reference point for dispute resolution than is the quality of trust between the organizations” (123). Different from all these studies, Klein Woolthuis, Hillebrand and Nooteboom (2005), through comparative case studies of four inter-firm relationships, find that the relationship between formal and informal contracting is so complex and dynamic that they can be both complements and substitutes, largely dependent on managerial contexts.

In public management literature, the empirical research on whether formal contracting and relational contracting are substitutes or complements is still less common. Van Slyke (2006) finds through interviews with public and nonprofit managers that social service contracting management might evolve from more formal-contracting like to more relational-contracting like over time. Lambright (2009) examines the use of government contracting monitoring tools from both

principal-agent perspective (formal contracting) and stewardship perspective (relational contracting). She concludes neither one could explain the entire story. Lamothe and Lamothe (2012) confirm this argument and find that in local service delivery, there are substantial contact and communicate between public managers and their vendors in contract implementation, in addition to clearly written formal contracts. Put together, the evidence so far points to the coexistence of these two mechanisms that public managers would devote themselves to simultaneously. To some extent, such combination of formal and informal contracting reflects the nature of contracting management in public administration context: well-planned and written contracts to meet the formal accountability demand, and negotiation and discretion to satisfy the flexibility concerns in service delivery (DeHoog, 1990).

In sum, this chapter provides a theoretical framework of contract design for human services. Holding a principal-agent perspective, this chapter first argues that formal contract design depends on two dimensions: task programmability and outcome measurability of the contracted services, which further lead to two contract arrangements: behavior-based contracts and outcome-based contracts. However, given that human services share both low task programmability (due to high provider discretion) and low outcome measurability (due to multidimensional, long-term outcomes), neither formal contract arrangements might fit seamlessly with human services. To provide a balanced theoretical framework, this chapter also includes the literature on relational contracting, which implies the reliance on relational exchange as an informal contracting management mechanism. Put together, the combination of

formal and informal contracting literature provides a complete framework to study contracting management. With this theoretical framework in mind, this project turns to the discussion of vocational rehabilitation, a human service area where PBC is becoming increasingly prevalent, as a policy field of inquiry for this present research. Particularly, Indiana vocational rehabilitation program's transition from FFS to PBC in the purchase of VR employment services provides a good case to answer the PBC effectiveness question raised earlier.

## Chapter 4. Vocational Rehabilitation as a Policy Field

### 4.1 Vocational Rehabilitation Programs

In the United States, 56.7 million American had a disability in 2010, representing 18.7 percent of the population. Among them, about 41 percent of those aged 21 to 64 with a disability were employed<sup>2</sup>. Employment has been found to be fundamental to people's physical and psychological well-being (Dooley, Fielding, & Levi, 1996; Linn, Sandifer, & Stein, 1985; Paul & Moser, 2009). Employment would help disabled people move toward desired quality-of-life changes. However, the disabled generally face a number of barriers in entering into the workforce and inclusion to the society. This calls for public vocational assistance. The major public vocational assistance service for adults with disabilities in the United States is the federal vocational rehabilitation program.

The federal interest in rehabilitation issues started in 1920s, with the enactment of the Vocational Rehabilitation Act, also known as the Smith-Fess Act. The Act begins the federal-state partnership in the rehabilitation of individuals with disabilities. The passage of the Rehabilitation Act in 1973 marked a significant progress in the federal rehabilitation program. It provides the statutory authority for programs and activities that help individuals with disabilities in the pursuit of gainful employment, independence, self-sufficiency, and full integration into community life. Under the Act, a wide range of rehabilitation programs were created. The U.S. Department of

---

<sup>2</sup> <http://www.census.gov/newsroom/releases/archives/miscellaneous/cb12-134.html>

Education has primary responsibility for administering the Act, particularly the programs under the Act that are funded through the Department of Education. Within the Education department, the Rehabilitation Services Administration (RSA) is the principal agency for carrying out most of programs and activities that provide direct support for vocational rehabilitation (VR), independent living, and individual advocacy and assistance.

By far, the largest program administered by RSA is the State Vocational Rehabilitation Services Program, also known as the Vocational Rehabilitation State Grants Program. Title I of the Rehabilitation Act of 1973 authorizes the VR program to “empower individuals with disabilities to maximize employment, economic self-sufficiency, independence, and inclusion and integration into society.” This program funds state VR agencies to provide employment-related services for individuals with disabilities to prepare for, gain, and maintain employment. The value of VR programs has been well recognized (Bolton, Bellini, & Brookings, 2000; Bond, 2004; Dutta, Gervey, Chan, Chou, & Ditchman, 2008; Gamble & Moore, 2003). Typically, the VR program service more than 1 million people with disabilities nationwide each year. More than 90% of the people who use state VR services have significant physical or mental disabilities that seriously limit one or functional capacities, such as mobility, communication, and interpersonal skills. The employment rates of people with disabilities after receiving VR services have been consistently found to be around 60% (Kaye, 1998).

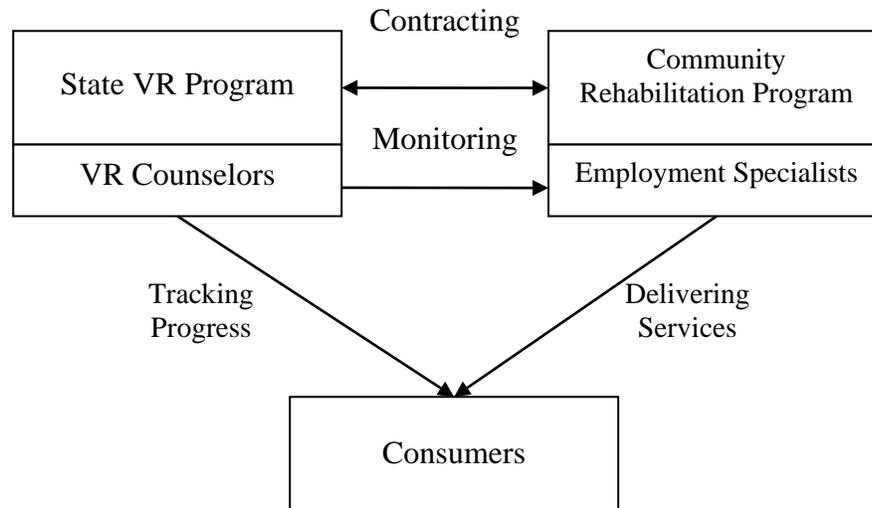
The VR program follows a federal-state model. Within the partnership, the federal government substantially funds state programs and states are also required to match federal funds. Generally, the federal government covers 78.7% of the program's costs through financial assistance to the states for program services and administration. For example, in fiscal year 2010, VR programs received \$3,040,323,049 federal funding and states expended \$ 864,073,243 (U.S. Department of Education, 2010). The federal government also establishes the program and monitors state program operation. For example, RSA conducts periodic on-site reviews and requires state VR agencies to submit annual program review, in order to ensure the state follow the program goals and requirement under the Rehabilitation Act. States enjoy certain latitude in running their VR programs and are responsible for delivers various VR services to clients. This federal-state vocational rehabilitation constitutes the policy field for the present research project.

#### 4.2 The Purchase of Job-related Services

As mentioned, the importance of employment for people with disabilities has been widely accepted. Thus, job placement and on-the-job support of people with disabilities at the highest level possible has been central to the mission of VR programs (Rubin, Roessler, & Dunkerby, 1983). Through these job-related services, VR programs help clients prepare for, gain, and maintain employment. Specially, the job-related services in VR include job search assistance, job placement assistance, and on-the-job support. Often, state VR agencies acquire these services from nonprofit community rehabilitation programs, through a variety of purchase of

service contracts. Figure 4 describes the general contractual relationship in the purchase of job-related services.

Figure 4. The Contractual Relationship in the Purchase of Job-Related Services



Three major players are involved in the rehabilitation process:

- Vocational rehabilitation counselor: The VR counselor is a rehabilitation professional, usually with a master degree level, who is an employee of the state VR program. The counselor is usually knowledgeable about consumers with disabilities and their vocational needs and thus determines the eligibility for VR services. The counselor is also responsible for assisting the consumer to determine and achieve a suitable vocational objective. The counselor works with the customer to devise an individual employment plan that will lead to the achievement of the vocational objectives. The counselor is responsible for authorizing service contractors for service needs, assuring the services

delivered are appropriate, and issuing payment based on service amount and consumer achievement.

- Contractor: a vendor of services, mostly a nonprofit community rehabilitation program, who has a contract with VR agency to deliver specific services leading to employment of consumer in a competitive job<sup>3</sup>. VR services, such as job placement assistance and on-the-job support, are generally delivered by an employment specialist, who directly works with a consumer. The VR counselor makes authorizations against their contract for specific services.
- Consumer: an individual with a disability who has been determined eligible for VR services by the VR counselor.

Traditionally, these service contracts are process-oriented, making contract compensation contingent upon the provision of services. Most of these contracts have common elements: defined services, a purchasable unit for each service (e.g., day, hour), and a unit cost for each defined service (Revell, West, & Cheng, 1998). The

---

<sup>3</sup> Competitive employment means work in the competitive labor market that is performed on a full-time or part-time basis in an integrated setting; and for which an individual is compensated at or above the minimum wage, but not less than the customary wage and level of benefits paid by the employer for the same or similar work performed by individuals who are not disabled. An Integrated setting is typically found in the community in which individuals interact with non-disabled individuals, other than support staff, to the same extent that non-disabled individuals in comparable positions interact with other persons.

predominant purchasable unit for services is an hour. For example, a contractor may be paid \$30 for each hour of job placement service it provides to an eligible service recipient. The popularity of hour-based contracts lies in several aspects. First, service contractors can customize service based on individual service needs, because they are reimbursed for each hour of service provided to individuals. Second, funding agencies have access to individualized information on the specific services provided and the impact of their funds. Through intensive reporting by service providers throughout the delivery process, funding agencies are able to control the services needed for successful employment and the detailed flow of funds. In that way, VR agencies actually centralize the service delivery process.

However, the weakness of this contracting method is visible. The hourly fee-for-service contracts do not readily encourage quality assessment and quality control by service providers, as the services are paid for without considering the results of those services. Moreover, contractors have limited incentives to encourage service recipients to move toward desired employment outcomes. In essence, the hour-based contracts emphasize the provision of service per se, i.e., the time spent providing those services, rather than the results of those services. Indeed, this contracting method equips contractors with disincentives to pursue valued outcomes (client independence). Basically, hourly billing tends to bear an inverse relationship to client independence: it is in contractors' fiscal interest to emphasize service provision and hours billed rather than working toward employment and long-term stability. This

demonstrates a “triumph of process over results” (Kettner & Martin, 1993, 62) and further leads to high service costs and poor employment outcomes.

Therefore, there was an incentive for a more effective contracting approach that simultaneously considers valued employment outcomes and the costs to achieve those outcomes. Inspired by the national performance movement, PBC emerged as a new approach in the purchase of VR placement services. Under PBC, contractors are compensated for the outcomes of services rather than the process of service delivery. Thus, the defining feature is payment for the valued accomplishments of service recipients. This transition from FFS to PBC aims to pay for meaningful and measurable employment outcomes at a defined cost. Contractors receive payment only if the service recipients they serve successfully achieves defined employment outcomes, such as assessment, obtaining employment, and job maintenance for a specific time period. For example, the provider may be reimbursed \$ 1,000 when a service recipient finds a job and \$ 1,500 when this client reaches stabilization on the job.

Table 10. Components of VR Services and Contract Type

Service Component			Contract type
Inputs	Resources	Staff, facilities, ...	Fee-for-service
Process	Program activities	Job assessment, development, coaching	Fee-for-service
Outputs	Service delivery	Completion of services	Performance-based
Outcome (short-term)	Benefits of services	Job placement, retention	Performance-based
Outcome (long-term)	Long-term quality-of-life changes	self-sufficiency, independence, and	Performance-based

		inclusion	
--	--	-----------	--

Source: Novak et al. (1999).

As Novak, Mank, Revell, and O'Brien (1999) argue, PBC in VR service promises a number of benefits: increased emphasis on valued outcomes and accountability for results, increased cost efficiency and effectiveness due to streamlined service delivery, and increased consumer choices and satisfaction. First, PBC approach compensates contractors when service recipients attain successful employment outcomes, rather than reimbursing the amount of services delivered and time spent. The success of services lie not in the array or number of services provided but in the extent to which these services embrace desired results. Along with the innovation, there is a change in the institutional environment of VR programs, from an accountability for following rules and regulations to an accountability for outcomes, in line with the government-wide performance movement. PBC thus enables contractors to increase accountability for aligning resources to achieve results.

Second, PBC promotes streamlined service delivery and improves cost efficiency and effectiveness. Under PBC, service providers are granted greater flexibility in service delivery in return for greater accountability for service performance. It deemphasizes regulations and micromanagement of contractor operation throughout the service process. Thus, time spent in reporting and paperwork would be largely squeezed. Such saving from documentation and reporting is supposed to be devoted to carefully

serving people with disabilities. This further encourages more cost efficient and effective service delivery.

Third, with an outcome orientation, contractors are expected to behave toward more effective service delivery. This will lead to the achievement of more timely outcomes for service recipients, and thus, increased customer satisfaction. In short, PBC is expected to generate a triple win for VR programs: disabled people receiving quick and quality services, contractors enjoying less regulation and greater flexibility, and state VR agencies achieving better results at lower costs with greater accountability (Frumkin, 2001; O'Brien & Revell, 2005).

#### 4.3 PBC Models in VR Services

##### *Oklahoma Milestone Payment System*

Oklahoma is a pioneer in the design and use of PBC in the purchase of VR services. The Oklahoma Department of Rehabilitation Services (DRS) began providing employment assessment and training services for people with severely mental and developmental disabilities in 1988, through contracting with community nonprofit service vendors. After receiving rehabilitation services, eligible individuals were able to achieve placement in local communities. Typically, these services were purchased from nonprofit contractors on a fee-for-service basis that reimburses nonprofit contractors at hourly payment rates for all services provided.

However, the DRS soon found the program experienced high costs but poor performance in helping the disabled for integrated employment in their communities. In 1991, bringing one case to closure cost more than \$22,000 and took 438 days on average (Frumkin, 2001). The DRS attributed this to the distorted incentives in the fee-for-service method: it emphasized contractor efforts in delivering services rather than in achieving employment outcomes through those services. This further led to an inverse relationship between contract payment based on amount of services provided and employment outcomes. To address the problem, the DRS designed the Milestone Payment System, in which contractors were reimbursed when service recipients reached each of milestones leading to employment and long-term stability.

The DRS defined each milestone as a predefined check point on the way to a desired outcome, such as case assessment, job placement, and job retention. Each milestone may include quality outcome indicators to be accomplished before payment. For example, consumer and employer satisfaction with job placement and minimum working wage were used as the quality indicators for the milestone on placement. Each milestone is associated with a fixed rate payment, with the higher payments toward the later milestones. The payment rate at each milestone would reflect the average cost of achieving the specific milestone rather than the cost of staff time (as under FFS model). Payment rates were negotiated for each milestone. The DRS solicited bids from community rehabilitation programs, allowing vendors to include in average cost per closures from the previous year multiplied by the estimated number of closures for the contract year. The DRS then reviewed the bids primarily

based on the per-customer bid price and the average cost per closure, as well as past service history. After that, the DRS negotiated with community vendors to achieve agreements (Frumkin, 2001). An example of milestone payment structure is: (1) determination of consumer needs – 10 % of bid; (2) vocational preparation completion – 10 % of bid; (3) job placement – 10% of bid; (4) 4 weeks job retention – 20% of bid; (5) job stabilization – 20% of bid; and (6) consumer rehabilitated (stabilization +90 days) – 30 % of bid.

The milestone payment system was first piloted in 1992. After several years’ pilot, the DRS converted all the service contracts to the milestone approach in 1997. Effectively July 1, 2001, the DRS moved the milestone payment system to the statewide fixed rates. Table 11 provides an example of the Oklahoma milestone payment system for the purchase of supported employment services<sup>4</sup>.

Table 11. Oklahoma Milestone Payment System

<b>Milestone</b>	<b>Regular Rate</b>	<b>Highly Challenged Rate</b>
Assessment and Career Planning	\$ 625	\$ 625
(Optional) Vocational Preparation	\$ 625	\$ 625
Job Placement	\$1,688	\$3,125

<sup>4</sup> Supported Employment Services is intended for individuals with the most significant barrier to employment who require: (1) substantial assistance in making a job choice, (2) substantial assistance in getting a job matching that choice, (3) a significant degree of job site support to learn the job tasks, gain work adjustment skills, and stabilize in employment, and (4) long-term support to retain employment.

4 Weeks Job Support	\$2,250	\$1,875
8 Weeks Job Support	\$1,688	\$1,875
Job Stabilization	\$2,125	\$2,125
Successful Employment	\$2,875	\$4,125

<b>Milestone</b>	<b>Outcome Description</b>
Assessment and Career Planning	A determination of the individual's informed job choice has been made, and the specific supports the individual will need to perform the chosen job successfully have been identified.
Vocational Preparation	The individual has clarified his/her career/employment objectives which include short-term and long-term vocational goals developed collaboratively with the individual.
Job Placement	The individual has been placed in a job of his/her choice meeting the requirements of supported employment and the objective in the IPE. An individual under this contract may not become an employee of the Contractor. Job placement is complete when the individual has completed the fifth day of work.
4 Weeks Job Support	The individual has worked successfully for a minimum of four weeks, beginning with the first day of employment (note 1).
8 Weeks Job Support	The individual has worked successfully for a minimum of 8 weeks total and has received the appropriate support services (Note 1).
Job Stabilization	The individual has worked successfully for the minimum required weeks (a total of 12 weeks for individuals receiving services under the regular rate and 17 weeks for individuals who are highly challenged) and is working the weekly work goal as identified in the IPE (Note 2).

Successful Employment	The individual has been employed a minimum of 90 days beyond stabilization and the case is ready for closure (note 2).
-----------------------	------------------------------------------------------------------------------------------------------------------------

Note 1: Only weeks in which the work hours exceed 40% of the weekly work goal, and in which on-site and/or off-site supports are provided, will be counted towards the minimum four weeks of this milestone.

Note 2: Only weeks in which hours worked meet the weekly work goal, and where needed supports were provided will be counted.

Source: Metro employment services contract 2012, Department of Rehabilitation Services, State of Oklahoma

Under the milestone contracting approach, service contractors were reimbursed when clients they served achieved certain milestones along the way to successful employment. The DRS did not specify the vocational methods to be used; vendors had the flexibility in achieving specified outcomes. To encourage contractors to take on more difficult clients, the milestone system designed a two-tiered payment rate, with a different rate for serving highly challenged clients. VR counselor, working with the individual and the contractor, designated the services to be used and whether the individual fit the regular or highly challenged rate. Services would be purchased on an individual basis as authorized by the counselor. Each milestone would be pre-authorized by the counselor and paid only once per case, per contractor, upon receipt and acceptance of the required documentation for payment by the counselor. Payment of a milestone would constitute payment in full for all services delivered during that phase of the program.

In short, the milestone payment system created different incentives for contractors.

Under the hourly payment system, the provider generated more income by delivering

more services before placing customers. Under the milestone payment system, the providers' incomes improved when consumers got jobs of their choices as rapidly as possible (O'Brien & Revell, 2005).

The Oklahoma PBC system received extensive recognition and was introduced by other states—including Alabama (Valerie, Howard, Dan, Byron, & Amy, 2000), Colorado (Block, Athens, & Brandenburg, 2002), Indiana (McGrew, Johannesen, Griss, Born, & Katuin, 2005), New York (Gates et al., 2004), etc. --into their purchase of VR services. Although there are small variations in the PBC systems across states, all these systems were modeled after Oklahoma.

#### *New York PBC Demonstration*

The New York State Office of Mental Health implemented a 2-year demonstration of PBC to promote employment outcomes (placement and retention) for people with serious mental health conditions, starting 2000 (Gates et al., 2004). Before the initiative, a traditional fee-for-service method was used, where providers were paid quarterly advances for hours spent working with clients regardless of consumer outcomes.

The demonstration model included 6 milestone payment points – life skills assessment, vocational planning and initial job placement, job skills acquisition, retention at 3 months, retention at 6 months, and retention at 9 months. Each

milestone was associated with a fixed rate payment, with higher payments toward later milestones. The rate was determined by the government agency, factoring in provider-estimated costs with provider-estimated consumer success rates at each milestone. Additional funding was available for long-term support, encouraging contractors to offer time-unlimited support to consumers once they had completed milestone VI (retention at 9 months). The same as other PBC models in VR, New York also developed a two-tiered payment to avoid creaming problem. Providers serving the most difficult clients would receive 20% more payment than serving the standard clients.

Table 12. New York State Milestone Payment System

<b>Milestone</b>	<b>Standard Rate</b>	<b>Incentive Eligible Rate</b>
Life Skills Assessment	\$ 750	\$ 900
Career Planning & Initial Job Placement	\$ 750	\$ 900
Job Skill Acquisition for 4 Weeks	\$1,500	\$1,800
Job Retention at 3 Months	\$1,500	\$1,800
Job Retention at 6 Months	\$1,875	\$2,250
Job Retention at 9 Months	\$1,125	\$1,350
Long-term Job Supports	\$1,300/year	\$1,300/year

Source: O'Brien and Revell (2006).

#### *Indiana Result-based Funding System*

To date, Indiana is the latest state that changed from a traditional fee-for-service model to a performance-based model, or what they call result-based funding (RBF).

The transition began with a pilot project. In 2002, The Indiana Supported Employment Results-based Funding Pilot Project was launched by the Supported Employment and Consultation Training (SECT), the Office of Vocational Rehabilitation Services (VRS), and the Indiana Division of Mental health and Addictions (DMHA) in supported employment services for individuals with severe mental health problems. Stakeholders, including government staff representatives, contractor, and consumers, were actively involved in the planning stage to determine the structure of the RBF system. The pilot RBF system included: (1) completion of the person-centered plan - \$550 (10% of VRS funding), (2) consumers' 5<sup>th</sup> day of employment - \$1,100 (20% VRS funding), (3) 1 month of employment - \$1,100 (20% VRS funding), (4) VRS eligible case closure - \$2,750 (50% VRS funding), and (5) 9 months of continuous employment - \$1,000 (DMHA funding) (McGrew, Johannesen, Griss, Born, & Katuin, 2005). The total amount paid for milestones 1-4 reflected the statewide historical average paid by VR per successful case closure under the FFS model, plus an amount equal to the average costs of providing services for individuals who fail to reach case closure. The milestone 5 was used as an extra bonus to incentivize long-term retention.

The Indiana Bureau of Rehabilitation Services changed its statewide contracting approach to RBF in late fiscal year 2006. Table 13 provides an example of its RBF system. The emphasis of RBF was placed upon structuring service contracting method that would increase the likelihood of both initial job placement and long-term tenure. Under RBF, contractors received reimbursement at a fixed rate once

consumers reached predetermined stages across the employment process. VR counselors would make the decisions on milestone authorization and the tiers individuals will enter. Each milestone should be authorized at the completion of the prior milestone. Providers should not provide services without proper authorization. Substantial progress towards the vocational goals needed to be demonstrated throughout the service process.

Table 13. Indiana Result-based Funding System

<b>Milestone</b>	<b>Tier I Rate</b> (For people who need ongoing support)	<b>Tier II Rate</b> (For people who do NOT need ongoing support)
1. Plan for Employment & Supports	\$1,200.00	\$ 600.00
2. Job Placement	\$1,200.00	\$ 900.00
3. Four Week Placement	\$1,864.00	\$1,325.00
4. Eligible for Closure	\$4,000.00	\$2,600.00
<b>TOTAL</b>	<b>\$8,264.00</b>	<b>\$5,425.00</b>

Note:

Tier One: For people who (1) qualifies as the most severely disabled as defined in the state policy, (2) requires multiple services over an extended period of time, and (3) is likely to need ongoing, intensive intervention to get and keep a job.

Tier Two: For people who (1) has a disability, severe disability, or most severe disability, and (2) would not require ongoing, intensive intervention to get and keep a job.

<b>Milestone</b>	<b>Outcome Description</b>
1. Plan for employment & supports	A plan for employment and supports developed by the customer and his/her support team. The team is

	comprised of the customer, Vocational Rehabilitation Counselor, employment service provider and any other stakeholder or individual the customer desires to participate in the meeting.
2. Job placement	The customer has worked one week at the hours per weekly work goal (e.g., based upon hours scheduled) in the vocational area identified in the Plan for Employment and Supports.
3. Four week placement	The customer has worked four weeks in which he/she met hours per weekly work goal (e.g., based upon hours scheduled) and pay rate as stated in the Plan for Employment and Supports. The customer is satisfied with the job. The employer has indicated satisfaction with the employee.
4. Eligible for Closure	The customer has maintained employment for 60 calendar days (for those eligible for Supported Employment services) or 90 calendar days for others. The customer is employed in a job as outlined in his/her Plan for Employment and Supports that is commensurate with his/her skills and abilities. The customer meets VRS closure criteria. Customer and employer are satisfied and this is documented (verbal or written reports).

Source: Indiana Bureau of Rehabilitation Services (2006).

#### 4.4 The Use of PBC in VR Services

As can be seen from the discussion above, the use of PBC in the purchase of job-related services has become widespread. More interestingly, unlike the PBC systems

used in other human service areas mentioned in chapter two, the PBC systems piloted and used in VR services in different states are roughly the same, most modeled after Oklahoma. This implies that the PBC in VR services has been relatively stable and mature, which provides a good policy field for this study to examine the question proposed in previous chapters – whether PBC is better than FFS.

Generally, the design of PBC system in VR employment services involves three common components: (1) defining the desired employment outcomes, (2) defining the payment point for each outcome, including criteria for determining achievement of each outcome, and (3) establishing a fee structure for payment points (Novak, Mank, Revell, & O'Brien, 1999). First, VR employment services generally proceed through several stages – establish job goal, become employed, stabilize in employment, and continue in employment. The design of the desired service outcomes is in line with these stages. Second, the selection of specific benchmarks and criteria to qualify a contractor for reimbursement needs to include consumer and employer satisfaction and the quality and stability of services. Third, the fee structure reflects the average cost of serving an individual within each defined outcome.

Indeed, the fee should include contractor costs associated with serving individuals who reach an employment outcome and costs historically associated with serving individuals who fail to achieve employment outcomes.

Actually, although some studies conclude promising results after the use of PBC, systematic evaluations of PBC effectiveness in VR services are still missing. For

example, after the use of the milestone payment system in Oklahoma, a study observes that customers' time on waiting list reduced by 53%, time before placement reduced by 18%, time from placement to success reduced by 45% observation, and number of people assessed but without placement reduced by 25%<sup>5</sup>. However, such before-after observation is not free from methodological problems. Further, another two problems related to PBC effectiveness are not well explored.

The first one is the customer selection or “creaming” problem. Due to the outcome orientation, contractors under PBC may prefer serving individuals who appear to be easier to place and thus have disincentives to serve people with the most significant disabilities. Contractors could maximize earnings by serving only the most readily employable people at the expense of serving those with more significant support needs. Although the PBC systems used in VR services all include two tiers of payment to consider the cost variance between serving regular customers and serving difficult customers, the effectiveness of such design is still largely unknown.

Second, the quality of employment outcomes deserves some more attention. Under PBC, contractors may be less likely to invest in job matches and job development to ensure quality employment. A good placement requires extensive services in job preparation and match for job seekers, which extends contractors' service time and cost. However, PBC fosters the achievement of employment milestones in a timely manner. In fact, the quality of these milestone achievement and the long-term effects

---

<sup>5</sup> [www.onenet.net/~home/milestone](http://www.onenet.net/~home/milestone).

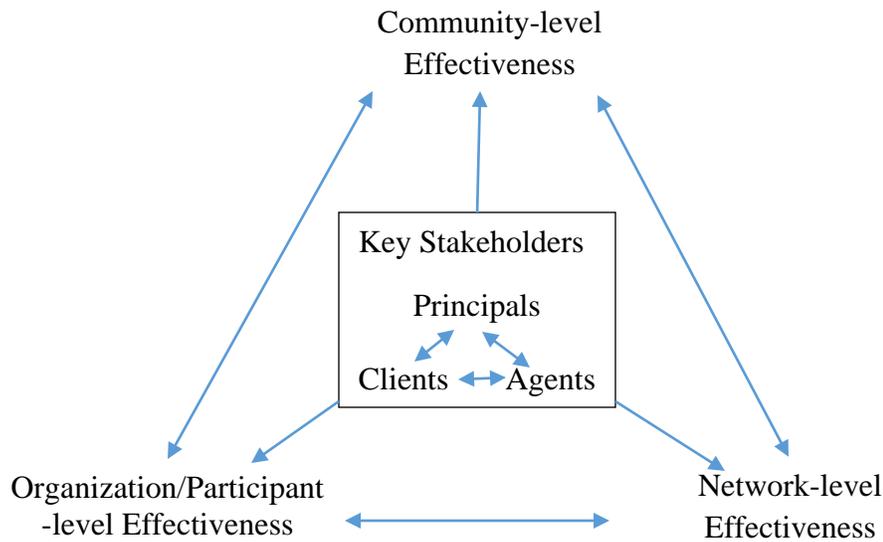
are hard to define and thus are not attached to milestone payment. Although some quality indicators such as customer satisfaction and working hours and wages at placement, they may not guarantee against contractors' gaming behaviors. More broadly, this questions the effectiveness of PBC on the aspects of employment outcomes that are not specifically measured in the PBC systems – do contractors gaming the PBC systems? Here, with these questions, we turn to the next two chapters that are devoted to examine PBC effectiveness, from both quantitative and qualitative perspectives.

## Chapter 5. The Effectiveness of PBC: A Service Outcome Perspective

### 5.1 Introduction

This chapter starts to evaluate the effectiveness of PBC. The performance of a contractual network, as Provan and Milward (2001) argue, is a multi-dimensional construct, including community, network, and participating organizations, each with different effectiveness criteria (Figure 5). At the community level, networks are evaluated by the contribution they bring to the communities and the clients they serve in addressing certain policy problems. In this way, the community perspective of network effectiveness means “first by assessing aggregate outcomes for the population of clients being serve by the network, and second, by examining the overall costs of treatment and service for that client group within a given community” (Provan & Milward, 2001, 417). At the network level, effectiveness should consider the operation of network structure per se. Thus, the effectiveness is evaluated based on the growth and function of the network as a whole, such as membership growth, range of services provided, and network maintenance. In addition, network effectiveness needs to recognize participating organizations involved, their individual survival and success in particular. This organization perspective would assess network on client outcomes, agency survival, legitimacy, resource acquisition, and cost.

Figure 5. Analytical Framework for Network Effectiveness



Source: Provan & Milward (2001).

Given the bilateral nature of government-contractor structure in PBC, with no network-level arrangement involved, this project assesses PBC effectiveness mostly from community-level and participating organization-level. This chapter holds a community-outcome perspective, leaving the other perspective to next chapter. Specifically, it explores whether PBC contributes to the improvement in client well-being. The research design is to quantitatively compare the employment outcomes under two contracting models, FFS and PBC. The unit of analysis is the individual client receiving placement services in Indiana. As mentioned, Indiana is the latest state in the transition from FFS to PBC, which provides a case that allows using administrative data to examine the policy impact of PBC intervention. Several approximations of employment outcomes are identified: likelihood of employment, time to placement, job retention, and wage. The first two are directly targeted by performance measurement in Indiana RBF system. PBC motivates contractors to

move through the performance milestones and across the employment process quickly in order to receive reimbursement. Thus, I predict:

H<sub>1</sub> After using PBC, clients are more likely to attain employment.

H<sub>2</sub> After using PBC, clients are able to achieve employment in less time.

In addition to these two indicators, another two employment quality indicators (job retention and wage), not directly targeted in the RBF system, are also included to examine if the potential performance improvement in employment possibility and time to placement is attained through gaming other unmeasured performance. As we may notice, these hard-to-measure performance areas are often excluded in performance measures in PBC systems. Even in Indiana RBF system, job retention (in terms of working hours) and working wages are measured by minimum standards such as state minimum hourly wage. Indeed, by leaving high discretion to contractors during the service process, PBC implicitly assumes nonprofit contractors would work with clients meticulously and innovatively and help them secure high-quality employment. Thus, this research would also test:

H<sub>3</sub> After using PBC, clients are able to achieve longer job retention.

H<sub>4</sub> After using PBC, clients are able to gain higher wages.

## 5.2 Research Design

This research uses an experimental design to examine the treatment effect of PBC in Indiana. Experimental designs are widespread in examining the treatment effects of policy interventions. The treatment effect in an experiment, intuitively, is the net difference between the condition of a unit after receiving a treatment and the condition of that unit if it would have not received that treatment. However, these two conditions are not possible to observe at the same time in real life: we can only observe one of these potential conditions, not both. This “missing data problem” (Rubin, 1976) constitutes what Holland (1986) called the “fundamental problem of causal inference” (947). Thus, the core task of policy/program evaluation study is constructing a counterfactual outcome to estimate the unobserved outcome. This implies that we have to find a control group that to a greatest extent approximates the treatment one in various aspects. To put it another way, ideally, there should be no selection bias between counterfactual outcome of treatment group and observed outcome of control group.

When the unit of analysis is a group and the analytical objective is average treatment effect, the composition of the samples in groups, or assignment mechanism used to assign samples to either treatment group or control group, becomes relevant. A key to estimate the causal effects is to identify a control that have units share the characteristics with those in treated group, or the distribution of covariates<sup>6</sup> is the

---

<sup>6</sup> Covariate is a variable that is measured before the treatment and thus is not affected by the treatment, such as many demographic variables.

same for treated and control groups. Generally, there are two experimental designs associated with two different assignment mechanisms, randomized experiments, in which units are assigned to different conditions randomly, and quasi-experiments, where units are assigned to conditions not by chance.

Admittedly, the randomized experimental design is the most desired design in evaluation studies. In this true experimental setting, the randomized control trial offers a robust and straightforward means to assess treatment effects. Random assignment guarantees that there would be no systematic preexisting differences between comparison groups before treatment and the only differences on all background covariates between two groups before the treatment, if any, are random, due to chance. This randomization removes selection bias, ensuring that all the characteristics of units are equally distributed between groups. Thus, the intervention is the only differentiating factor between groups. The average difference observed in outcomes can be attributed to the impact of the intervention (or/and possibly sampling error if the sample is not large enough).

However, for various ethical and practical reasons, this ideal randomized experimentation is not feasible in current research. Instead, I resort to quasi-experiment design. In quasi-experimental setting, samples are “collected through the observation of systems as they operate in normal practice without any intentions implemented by randomized assignment rule” (Rubin, 1997, 757). In this way, assignment of conditions is determined by factors beyond the experimenter’s control,

such as self-selection and administrator selection (Shadish, Cook, & Campbell, 2002, 13-14). Thus, it is very likely that “the treated and control groups differ prior to treatment in ways that matter for the outcomes under study” (Rosenbaum, 2002, 71). In this case, systematic difference in the characteristics besides the treatment may influence outcomes, and by directly comparing the difference between outcomes may fail to provide robust answers. Thus, more complicated research designs are needed to address this problem.

### 5.3 Interrupted Time Series with a Control Group Design

Although quasi-experiments, if well designed, are able to reproduce the same results as randomized experiments, extra attention should be paid to correct the potential challenges. This study uses an interrupted time series with a nonequivalent control group design, diagrammed in Figure 6, to examine the treatment effect of performance-based contracting. It compares individual employment outcomes in Indiana VR programs before and after the PBC intervention within a time period of 2004-2009. As mentioned, Indiana, as the treatment group, changed the purchase of VR placement services from FFS to PBC in the end of fiscal year 2006. Michigan, Indiana’s only neighbor state which kept fully using FFS over time, is added as a control group. The repeated cross-sectional data for analysis was requested from the Rehabilitation Services Administration (RSA) of the Department of Education. The RSA 911 database reports records pertain to all the individuals whose case records were closed in a given fiscal year, including personal characteristics, types of

services, and employment outcome of all clients receiving state VR services.

Figure 6. Interrupted Time Series with a Nonequivalent Control Group Design

	2004	2005	2006	2007	2008	2009
IN	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	X	O <sub>5</sub>	O <sub>6</sub>
MI	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O <sub>4</sub>	O <sub>5</sub>	O <sub>6</sub>

Here, the research relies on Campbell and Stanley’s (1963) typology on the threats to internal validity in quasi-experimental designs. The type of quasi-experimental design used here is robust in removing most of the threats to internal validity, such as maturation, testing, and regression, but still subject to instrumentation, local history, and selection (Cook & Campbell, 1979; Shadish et al., 2002). Thus, in order to gain a more accurate estimate of the treatment effect, these three potential threats should be minimized as much as possible before comparing treatment and control groups. Instrumentation may bias causal inference when different administrative procedures and measures are used to record participants’ performance over time. However, this would not be a big concern for state VR programs. Under the Rehabilitation Act, all the administrative and service components, procedures, and standards are under rigorous federal regulations. For example, RSA conducts annual reviews and periodic on-site monitoring of state VR programs to ensure they comply with program and performance requirements under the Rehabilitation Act. In this way, the consistency within and between states can be expected.

A serious threat comes from selection bias, i.e., differences exist between individuals in treatment and control groups. To solve this problem, matched sampling is used to correct the observed imbalances between the two states. Matched sampling is a resampling strategy, “selecting units from a large reservoir of potential controls to produce a control group of modest size that is similar to a treated group with respect to the distribution of observed covariates” (Rosenbaum & Rubin, 1985, 33). After matching, two comparison groups are identical on a variety of observed variables, which actually replicates a randomized experiment where the treatment assignment is unconfounded, at least given the observed covariates (Rosenbaum & Rubin, 1983; Rubin, 1973). In particular, this study adopts propensity score matching to produce the matched sample. A propensity score, as Rosenbaum and Rubin (1983) define, is “the conditional probability of assignment to a particular treatment given a vector of observed covariates” (41)<sup>7</sup>. Matching samples based on propensity scores allows simultaneously considering a variety of covariates. Rather than requiring exact or close matching on all covariates separately, propensity score matching enables matching on the scalar summary of the covariates. Given the propensity score, each unit has the same chance to be assigned to treatment, as in a randomized experiment. In essence, the propensity score is a balancing score. Given a propensity score,  $e(x)$ , the distribution of the observed covariates  $x$  is the same in both treatment and control groups. Matching treated and control based on the propensity score could create new comparison groups that are identical on a vector of those observed covariates

---

<sup>7</sup> The propensity score for subject  $i$  ( $i=1, \dots, N$ ) to be assigned to treatment ( $Z=1$ ) versus control ( $Z=0$ ) given a vector of observed covariates  $x_i$  is  $e(x_i) = \Pr(Z_i=1 | x_i)$

(homogeneous)<sup>8</sup>, replicating a randomized experiment based on these covariates.

Rosenbaum and Rubin (1983) also proved that treatment assignment and the observed covariates are conditionally independent given the propensity score. Thus, exact matching based on propensity score enables to remove bias due to all observed covariates and to produce unbiased estimation of the average treatment effect, measured by the difference in means in the outcome between treated and control groups.

Local context might also bias causal inference when the individuals in comparison groups reside in different settings. To address this issue, this study chooses Michigan as the control group against Indiana, aiming to maximize the socio-economic similarities between the two. In addition, I use difference-in-differences (DID) regressions after matched sampling to further adjust the unobserved imbalance. Under the DID model, any bias caused by exogenous variables common to Indiana and Michigan could implicitly be controlled for, even when these variables are unobserved. There is indeed some evidence to support the common trend assumption of the DID model during 2004-2009. The state-level factors that might affect employment outcomes, including GDP growth, unemployment rate, average weekly earnings, and VR program capacity (measured by average number of clients served per program staff) were found to roughly follow the same trend (See Appendix 1). I also reviewed the annual review reports of Indiana and Michigan VR programs and

---

<sup>8</sup> It is possible that two units with the same propensity score may be different in a certain observed covariate, but those differences are not systematic (Guo & Fraser, 2010).

didn't find any major policy changes on the purchase of employment services.

Therefore, I have somewhat strong confidence in assuming that the two states have parallel trends over time.

Indeed, running DID regressions on matched samples embraces a number of advantages. First, the combination of the two methods is most robust and efficient in removing the biases due to covariates and estimating the treatment effect on the treated (Abadie & Imbens, 2006; Heckman, Ichimura, & Todd, 1997; Rubin, 1973; 1979). A major problem in the use of matched sampling is inexact matching—it is not always possible to find enough matched treatment and control samples with exactly the same observed characteristics (Rubin, 1979). This is especially the case as the number of matching variables increases. Given the imperfect matching, the estimated treatment effect might not be accurate. However, when putting matched sampling and model-based regression together, matched sampling substantially reduces observed covariate differences, and model-based adjustment afterwards could further controls for residual differences.

Second, matched sampling relaxes the DID identification restrictions. Model-based regression adjusts the effect of confounding variables by estimating the relationship between the dependent variable and the confounding variables. The major problem associated with this method is that the model assumptions may be unwarranted in many cases. For example, the linear relationships with the dependent variable and matching variables may not be justified (Rubin, 1979). Thus, the combined method

makes model-based adjustment less sensitive to model specification. This again allows the estimation of parsimonious parametric approximations of the average treatment effect on the treated. (Abadie, 2005; Ho, Imai, King, & Stuart, 2007).

Guo and Fraser(2010), through a data simulation, also show that under the ideal conditions like randomized experiment, both methods work equally well, leading to accurate estimation of treatment effect with biases closing to zero. However, in quasi-experimental situation, especially when treatment assignment is not ignorable, although either method could remove the biases to a different extent, neither method could produce unbiased estimation of treatment effect. Also through a simulation, Rubin (1973; 1979) finds that model-based adjustment could produce smaller standard errors than matched samples when the model is correctly specified.

However, when the model is inaccurate, model-based adjustment would be less robust, not remove biases, but increasing them. Given this, it is suggested that the combination of matched sampling and regression adjustment to be the most robust method for producing the least biased estimate and controlling the biases due to the imbalances in observed covariates (Cochran & Rubin, 1973; Rubin, 1979). In a word, as Abadie (2005) suggests, this combined method “allow[s] for the distribution of both observed and unobserved factors to differ between treated and untreated, as long as the effect of unobserved factors on the outcome does not vary with time (or, more generally, if it experiences the same variation, on average, for treated and untreated)”

(5).

## 5.4 Propensity Score Matching

Propensity score matching was first used to produce matched samples. When conducting propensity score matching, I followed the procedures suggested by Caliendo and Kopeinig (2008) and Guo and Fraser (2010).

### *1. Specification of Conditioning Model*

The first step in conducting propensity score analysis is determining which covariates and conditioning model to be used to estimate propensity score. After all, the accuracy of the specification of covariates and models would affect the effectiveness of propensity score analysis and final estimation of the treatment effect (Heckman, Ichimura, & Todd, 1997; Rubin, 1997). However, there is no guideline available in current literature on propensity score analysis providing definitive answers.

Theoretically, in order to meet the assumption of ignorable treatment assignment, all covariates that might be related to treatment assignment and the outcome should be included into the conditioning model (Glazerman, Levy, & Myers, 2003; Rubin & Thomas, 1996; Stuart & Rubin, 2007). Omitting important variables would seriously increase bias in resulting estimates (Dehejia & Wahba, 1999). Shadish et al. (2008) warn that only relying on small set of “predictors of convenience,” such as demographic factors, would lead to poor matching performance. However, in most cases there is no comprehensive list of such conditioning variables explicitly. Therefore, to satisfy the assumption of strong ignorability to a great extent, scholars

generally suggest including a large set of covariates of theoretical relevance (Greevy, Lu, Silber, & Rosenbaum, 2004; Lunceford & Davidian, 2004). Actually, including variables that are little unassociated with the outcome might slightly increase variance, but excluding potentially important variables would increase bias. As Rubin and Thomas (1996) argue, “unless a variable can be excluded because there is a consensus that it is unrelated to outcome or is not a proper covariate, it is advisable to include it in the propensity score model even if it is not statistically significant” (253). This paper follows this convention.

In view of theoretical relevance and data availability, in this study, we include three categories of covariates in Table 14 —demographic background (age, education, race, gender, veteran status, primary disability, secondary disability), pre-service status (employment status, work disincentives, previous service status, Projects with Industry status), and employment service received (number of placement services received)—that are thought to be related to either treatment assignment or outcome.

Table 14. Description of Matching Variables

<b>Matching variables</b>	<b>Description and Measurement</b>
<i>Demographic Background</i>	
Age	An individual’s age at service application
Education	An individual’s level of education attained at application, with 0=less than high school, 1=special education, 2=high school graduate, 3=post-secondary/associate degree, and 4=college degree or higher
Race	An individual’s race and ethnicity, with 0=black or African American, 1=Native American (American Indian, Alaska native, native Hawaiian, or other Pacific islander), 2=Asian,

	3=white, 4=Hispanic or Latino
Gender	An individual's gender status, with 0=male, 1=female
Veteran	An individual's veteran status, 0=not a veteran, 1=veteran
Primary disability	An individual's primary physical or mental impairment, with 0=sensory/communication impairments, 1=physical impairments, 2=mental impairments
Secondary disability	An individual's second physical or mental impairment, with 0=no impairment, 1=sensory/communication impairments, 2=physical impairments, 3=mental impairments
<b><i>Pre-service Status</i></b>	
Employment status	An individual's employment status at application, with 0=not employed, 1=employment
Work disincentives	The number of public support an individual had at application, including supplemental security income (SSI), Temporary Assistance for Needy Families (TANF), general assistance from state or local government, social security disability insurance (SSDI), veterans' disability benefits, workers' compensation, Medicaid, Medicare, medical insurance not through employment, and others
Previous service status	If an individual had received previous employment service, with 0=no previous closure, 1= closed before services, 2=closed after services
Participation in Projects with industry	If an individual participates in Projects with Industry program, with 0=no, 1=yes
<b><i>Employment Services</i></b>	
No. of placement services received	The number of employment services an individual received throughout service process, including job search assistance, job placement assistance, and on-the-job supports

The propensity score, in its essence, is a balancing score representing a vector of covariates. Unlike in randomized experiments<sup>9</sup>, propensity scores in quasi-experiments are unknown and must be estimated. Propensity scores are often estimated using binary logistic regression with observed covariates X as independent variables and treatment assignment D (D=1 for treatment condition, D=0 for control

---

<sup>9</sup> In randomized experiments, each unit has a 50% probability of being assigned to either treatment or control group. Thus, the propensity score for each unit is 0.5, without considering sampling error.

condition) as dependent variable. The propensity score for unit  $i$  ( $i = 1, 2, \dots, N$ ) is as follows:

$$P(D_i|X_i = x_i) = E(D_i) = \frac{e^{x_i\beta_i}}{1 + e^{x_i\beta_i}} = \frac{1}{1 + e^{-x_i\beta_i}}$$

$\beta_i$  is the regression parameters

In using logistic regression to predict the value of propensity score, the aim of the modeling is not to estimate the parameters, but to balance the covariates between treatment and control groups. Thus, many traditional regression diagnosis methods, such as collinearity check and model fit statistics, are no longer helpful in model specification here. Rather, the balancing property of the propensity score is used to justify a model specification (Dehejia & Wahba, 1999; Rosenbaum & Rubin, 1984).

Researchers find that treatment effect estimation is not sensitive to the model specification used to predict the propensity score, as long as the balancing property of the propensity score holds (Waernbaum, 2010; Zhao, 2008). Misspecification of propensity score model under this condition would not bring bias to the treatment effect estimation.

The present project adopts the strategy suggested by Dehejia and Wahba (1999) and Rosenbaum and Rubin (1984), by correcting function form of covariates and adding higher order terms and interaction terms of observed covariates sequentially and

check the balance of the covariates based on the propensity scores. Particularly, I used STATA program `pscore.ado` developed by (Becker & Ichino, 2002) to estimate propensity scores. In particular, this program helps ensure the balancing property of propensity scores, i.e. observations with the same propensity scores should have the same distribution of observed characteristics, regardless of treatment status. The program first splits the sample into several spaced intervals of the propensity score and test whether the mean propensity score of the treatment and control units are statistically different within each interval. If the test fails in one interval, the program would split the interval in half and retest within each finer interval until the mean propensity score of the treatment and control units become balanced. Again, within each interval, testing the means of each covariate to ensure that there is no statistical difference between treatment and control units. If one or more covariates are not balanced in all intervals, the balance property is not supported by the current model specification and specification modification is necessary by adding more interaction and higher order terms.

## *2. Choose Matching Algorithm*

After estimating propensity scores, we move onto match treatment units with control units based on the value of propensity score. To date, there are already a number of matching algorithms available, including greedy matching, kernel matching, etc. Dehejia and Wahba (2002) highlight three major decisions in the choice of matching methods: (1) which matching algorithm to be used, (2) number of control units used to

match with each treatment unit, and (3) match with replacement or without replacement: reduce bias (with replacement); increase precision (without replacement); inexact matching vs. incomplete matching. Unfortunately, there has been no clear rule for determining which matching algorithm works best under what conditions; there is always a tradeoff between bias and efficiency. It is largely dependent on data per se and the research design.

In this study, we use 1-to-1 nearest neighbor matching within caliper without replacement, one of the most common matching algorithms, so-called greedy matching<sup>10</sup>. This matching algorithm randomly orders the treatment and control units, and selects for each treatment unit a control unit with the smallest distance from the treatment one. Once a control unit was matched to a treatment unit, it was removed from the control group without replacement. The most attractive feature of this matching algorithm is that it allows multivariate analysis used directly after matched sampling, without extra statistical adjustment (Guo & Fraser, 2010).

One limitation of this 1-to-1 nearest neighbor matching is that there is no restriction on the distance between two matched units, as long as they are nearest neighbors based on propensity scores. It is possible that these two units are very different in

---

<sup>10</sup> The trade-off between bias and variance: Matching one nearest neighbor minimizes bias at the cost of larger variance; matching using additional nearest neighbors increase the bias but decreases the variance. Matching with replacement keeps bias low at the cost of larger variance; matching without replacement keeps variance low at the cost of potential bias.

terms of propensity scores, but there is no one that is closer. Thus, I add a caliper (a quarter of a standard deviation of the propensity scores of the sample (Rosenbaum & Rubin, 1985) to the nearest neighbor matching, choosing matched units only when the absolute distance between the two units (in terms of propensity scores) are within a predetermined caliper.

The detailed algorithm is as follows: let  $p_i$  and  $p_j$  are the propensity scores for the treatment and control units  $i$  and  $j$ ,  $C(i)$  is the set of control units matched to the treatment unit, and  $\epsilon$  is the caliper. One control unit with the estimated propensity scores falling within a caliper  $\epsilon$  from  $p_i$  are matched to the treated unit  $i$ . The matched sample sets are:

$$C(i) = \{p_j \mid \| p_i - p_j \| < \epsilon\}$$

### 3. *Balancing Tests*

After matching, it is expected that the preexisted statistical differences in the covariate means between two comparison groups should be eliminated. And the two groups are comparable in that the distributions of observed covariates are identical in treated and control groups. Before moving forward, we need to check covariate balance before and after matching to ensure that covariate balance has actually been achieved.

I check covariate balance before and after matching using the absolute standardized difference in covariate means (D'Agostino, 1998; Haviland, Nagin, & Rosenbaum, 2007). The absolute standardized difference is the absolute value of the mean difference as a percentage of the average standard deviation. For each covariate  $X$ ,  $\bar{x}_t$  and  $\bar{x}_c$  are the means in the treatment and control groups, and  $s_t^2$  and  $s_c^2$  are the corresponding variances, respectively, the absolute standardized difference includes two standardized measures:

$d_x$  contrasts covariate values for treatment units with covariate values of all the potential controls before matching

$$d_x = \frac{|\bar{x}_t - \bar{x}_c|}{\sqrt{(s_t^2 + s_c^2)/2}}$$

$d_{xm}$  contrasts covariate values for treatment units with covariate values of all the matched controls after matching (a subscript  $m$  for before matching)

$$d_{xm} = \frac{|\bar{x}_{tm} - \bar{x}_{cm}|}{\sqrt{(s_{tm}^2 + s_{cm}^2)/2}}$$

Table 15 show the results of covariate balance check. For each year, the absolute standardized difference compares the covariate values of the treatment individuals with those of the control individuals before matching ( $d_x$ ) and with those of the matched control individuals after matching ( $d_{xm}$ ). T-tests examine the equality of covariate means in the treatment and control groups, both before and after matching.

Table 15. Covariate Balance Check Before and After Matching

(For individuals receiving employment services)

Covariate	2004			2005			2006			2007			2008			2009		
	$d_x$ (%)	$d_{xm}$ (%)	t-statistic	$d_x$ (%)	$d_{xm}$ (%)	t-statistic	$d_x$ (%)	$d_{xm}$ (%)	t-statistic	$d_x$ (%)	$d_{xm}$ (%)	t-statistic	$d_x$ (%)	$d_{xm}$ (%)	t-statistic	$d_x$ (%)	$d_{xm}$ (%)	t-statistic
	Before matching: N <sub>IN</sub> =2951, N <sub>MI</sub> =2148 After matching: N <sub>IN</sub> =N <sub>MI</sub> =1598			Before matching: N <sub>IN</sub> =3048, N <sub>MI</sub> =1143 After matching: N <sub>IN</sub> =N <sub>MI</sub> =955			Before matching: N <sub>IN</sub> =2673, N <sub>MI</sub> =1035 After matching: N <sub>IN</sub> =N <sub>MI</sub> =852			Before matching: N <sub>IN</sub> =2770, N <sub>MI</sub> =1213 After matching: N <sub>IN</sub> =N <sub>MI</sub> =1026			Before matching: N <sub>IN</sub> =2762, N <sub>MI</sub> =1098 After matching: N <sub>IN</sub> =N <sub>MI</sub> =970			Before matching: N <sub>IN</sub> =2569, N <sub>MI</sub> =887 After matching: N <sub>IN</sub> =N <sub>MI</sub> =785		
Age	22.6	12.1**	3.39	7.0	14.4**	3.14	11.8	17.7**	3.65	15.1	5.8	1.67	22.6	7.8	1.63	30.0	11.9**	5.61
Gender	4.0	6.9	1.93	7.4	5.8	1.27	5.5	10.6**	2.17	11.3	12.2**	2.75	7.8	13.3**	2.92	2.1	0.5	0.10
Race	23.7	1.8	0.50	6.7	6.1	1.30	12.7	7.2	1.46	12.5	5.4	1.63	17.0	2.0	0.40	14.0	5.8	1.07
Education	0.1	5.0	1.39	2.8	0.6	0.13	7.1	4.5	1.22	0.5	2.9	0.63	3.4	5.3	1.11	7.4	7.6	1.50
Veteran status	22.1	15.5**	4.44	18.4	20.7**	4.25	11.5	18.4**	3.41	6.8	7.9	1.70	20.9	7.8**	3.99	14.7	8.4	2.04
Projects with industry	2.9	2.8	0.71	7.7	1.9	0.45	4.6	0.0	0.00	8.9	1.9	0.58	2.8	2.3	0.38	9.8	2.1	0.45
Primary disability	12.0	5.4	1.47	18.7	16.7**	3.59	10.7	7.3	1.53	15.1	11.4**	3.29	15	7.7**	2.15	15.3	3.4	0.88
Secondary disability	36.6	6.0	1.69	37.2	16.3*	3.53	22.9	13.3**	2.75	21.7	16.6	3.80	28.7	13.6**	2.99	25.1	9.8	1.94
Employment status	10.8	3.5	0.94	4.4	0.0	0.00	7.8	6.5	1.33	1.0	2.8	0.63	9.2	1.7	0.48	0.8	1.2	0.23
Work disincentives	1.6	0.1	0.02	5.9	1.4	0.30	11.0	4.4	0.90	0.2	8.9	1.96	4.4	4.8	1.02	0.5	2.7	0.54
Previous closure/service	11.2	3.9	1.10	8.8	9.4**	2.03	8.7	9.0	1.86	6.5	3.5	1.00	16.3	4.5	1.25	13.2	6.2	1.59
No. of placement services received	14.9	8.0**	2.40	69.9	6.5	1.49	69.9	3.0	0.67	60.9	1.8	0.44	55.5	4.1	0.93	57.2	2.7	0.56

\*\*significant at .05; two-tailed tests.

**(For individuals with employment)**

	2004 Before matching: N <sub>IN</sub> =1185, N <sub>MI</sub> =862 After matching: N <sub>IN</sub> =N <sub>MI</sub> =525			2005 Before matching: N <sub>IN</sub> =1196, N <sub>MI</sub> =611 After matching: N <sub>IN</sub> =N <sub>MI</sub> =431			2006 Before matching: N <sub>IN</sub> =1303, N <sub>MI</sub> =553 After matching: N <sub>IN</sub> =N <sub>MI</sub> =376			2007 Before matching: N <sub>IN</sub> =1429, N <sub>MI</sub> =616 After matching: N <sub>IN</sub> =N <sub>MI</sub> =445			2008 Before matching: N <sub>IN</sub> =1277, N <sub>MI</sub> =550 After matching: N <sub>IN</sub> =N <sub>MI</sub> =398			2009 Before matching: N <sub>IN</sub> =1048, N <sub>MI</sub> =404 After matching: N <sub>IN</sub> =N <sub>MI</sub> =295		
Covariate	$d_x$ (%)	$d_{xm}$ (%)	t- statis tic															
Age	21.1	8.4	1.38	22.5	8.7	1.76	23.1	8.6	1.69	9.7	0.9	0.19	36.4	15.3**	2.97	25.0	17.8**	2.98
Gender	0.5	5.5	0.90	7.5	7.6	1.14	10.1	5.0	0.68	14.7	14.9**	2.20	11.0	3.6	0.71	4.0	5.6	0.68
Race	15.8	5.9	0.90	11.0	8.7	1.80	6.0	1.0	0.14	16.3	3.6	0.51	13.6	10.4	1.41	8.6	2.0	0.22
Education	3.7	4.7	0.73	8.9	6.0	1.39	22.2	7.0	1.34	11.7	5.1	1.05	13.2	12.7	1.70	13.4	4.6	0.54
Veteran status	25.2	12.5**	2.43	16.3	13.6**	2.96	44.1	18.2**	3.25	19.2	12.0**	2.33	4.9	3.9	0.45	7.0	9.1	1.08
Projects with industry	6.0	5.8	1.00	5.7	4.0	0.58	3.9	0.0	--	5.7	5.1	0.58	8.6	0.9	0.17	12.4	4.0	0.58
Primary disability	17.6	14.0**	2.23	23.2	18.8**	2.7	15.4	7.9	1.55	10.4	10.3	1.93	20.4	15.8**	3.03	8.2	1.1	0.19
Secondary disability	35.3	13.8**	2.21	31.1	12.9	1.91	23.9	18.3**	3.63	21.4	23.1**	3.51	25.9	7.2	1.35	33.4	15.5	1.89
Employment status	9.6	5.5	0.88	2.9	1.2	0.16	4.8	1.1	0.22	11.0	2.0	0.04	5.4	0.2	0.04	7.3	1.4	0.24
Work disincentives	0.5	1.7	0.27	12.0	7.4	1.08	17.7	13.8**	3.10	5.9	1.0	0.15	21.1	4.2	0.82	19.5	7.1	0.85
Previous closure/service	5.5	3.0	0.48	9.4	0.7	0.15	13.9	6.3	1.24	12.4	2.8	0.58	19.7	5.8	1.14	3.3	5.6	0.70
No. of placement services received	21.2	1.7	0.34	89.8	3.1	0.52	98.2	0.0	0.00	89.4	1.2	0.21	88.9	1.3	0.22	90.7	6.2	0.83

\*\*significant at .05; two-tailed tests.

For good matching,  $d_{xm}$  should be less than 5% after matching and t-statistic should be not significant after matching. In this vein, as can be seen from Table 15, the matched sampling in this study is quite effective in removing substantial part of the preexisting differences between two comparison states, but not all of them, as expected.

### 5.5 Difference-in-difference Regressions

After propensity score matching, matched sample has removed most of the imbalance between comparison groups at least in observed covariates. With the matched sample, I moved on to DID analyses to estimate the impact of PBC on Indiana clients, in terms of four employment outcome indicators. The general DID model is as follows:

For the logistic model on employment probability:

$$P(Y = 1) = F[\beta_0 + \beta_1(\text{Indiana}) + \beta_2(\text{Service Year 2007} - 2009) + \beta_3(\text{Indiana} * \text{Service Year 2007} - 2009) + \beta_4 X_1 + \beta_5 X_2 + \beta_6 X_3]$$

For the OLS models on time to placement, weekly working hours, and weekly earnings:

$$Y = \beta_0 + \beta_1(\text{Indiana}) + \beta_2(\text{Service Year 2007} - 2009) + \beta_3(\text{Indiana} * \text{Service Year 2007} - 2009) + \beta_4 X_1 + \beta_5 X_2 + \beta_6 X_3 + \mu$$

X1 contains “demographic background” variables, including age, education, race, gender, primary disability, and secondary disability.

X2 contains “pre-service status” variables, including employment status, work disincentives, previous service status, and participation in projects with Industry.

X3 contains “employment services” variable, i.e., number of placement services received.

Tables 16 and 17 present the DID regression results. Within each model, the interaction effect between the variable of Indiana and the variable of service period 2007-2009 is the differences-in-differences estimator of the treatment effect. First, logistic regression was employed to predict the differences in the likelihood of attaining employment result for those who received employment services before and after PBC. Before discussing the parameters of detailed variables, tests of goodness of fit of the regression model were performed. The logistic regression model is statistically significant (likelihood ratio chi-square=1102.74,  $p = .0000$ ), meaning that the model specified is significantly better than the model with only the constant. Hosmer-Lemeshow test for overall goodness of fit was also added. The Hosmer-Lemeshow chi-square equals to 8.943 ( $p = .063$ ), implying that the differences between the observed and fitted values are small. Both tests show that the logistic model is reliable to produce meaningful inference. Generally, after the introduction of

Table 16 Logistic Regression Model Predicting Likelihood of Employment for Service Recipients (N = 12, 372)

<b>Variable</b>	<b>Odds Ratio</b>	<b>Standard Error</b>	<b>Z Value</b>
<b><i>State and Service Year</i></b>			
State (Indiana)	0.7561***	0.0397	-5.33
Service Year 2007-2009	0.8287***	0.0450	-3.46
Indiana Service Year 2007-2009	1.4991***	0.1144	5.30
<b><i>Demographic Background</i></b>			
Age	0.9975	0.0016	-1.55
Education			
Special education	1.3384***	0.1207	3.23
High school graduate	1.2281**	0.1099	2.30
Post-secondary/associate degree	1.2460**	0.1249	2.19
College degree or higher	1.4369***	0.1783	2.92
Race			
Native American	0.4991**	0.1455	-2.38
Asian	1.4769*	0.3241	1.78
White	1.3715***	0.0707	6.13
Hispanic or Latino	1.4601***	0.1847	2.99
Gender (Female)	0.8292***	0.0326	-4.76
Veteran	0.8319*	0.0846	-1.81
Primary disability			
Physical impairments	0.7305***	0.0810	-2.83
Mental impairments	1.0239	0.1065	0.23
Secondary disability			
Sensory/communication impairments	1.0099	0.0.1084	0.09
Physical impairments	0.8492***	0.0480	-2.89
Mental impairments	0.8156***	0.0358	-4.46

<b><i>Pre-service Status</i></b>			
Currently employed	2.0271***	0.1071	13.38
Work disincentives	0.9278***	0.0160	-4.36
<b><i>Previous closure/service</i></b>			
Closed before services	0.8738*	0.0650	-1.81
Closed after services	1.2990***	0.0640	5.31
Participation in Projects with Industry	3.4439***	1.4556	2.93
<b><i>Employment Services</i></b>			
No. of placement services received	2.8868***	0.1392	21.99
<b>Likelihood ratio chi square</b>			
Pseudo $R^2$	.2653		

\*significant at .1; \*\*significant at .05; \*\*\*significant at .01; two-tailed tests.

Table 17. OLS Regression Models Analyzing Employment Outcomes (N = 4, 940)

Variable	Model (1) Time to placement			Model (2) Weekly working hours			Model (3) Weekly earnings		
	Coefficient	Robust Standard Error	t-value	Coefficient	Robust Standard Error	t-value	Coefficient	Robust Standard Error	t-value
<i>State and Service Year</i>									
State (Indiana)	123.7927***	9.69203	12.77	-2.7736***	0.3925	-7.07	-40.4802***	5.4046	-7.49
Service Year 2007-2009	28.1780***	9.7877	2.88	-1.2952***	0.3868	-3.35	2.3857	5.4851	0.43
Indiana Service Year 2007-2009	-72.1985***	14.3650	-5.03	1.3291**	0.5603	2.37	4.3715	7.2722	0.60
<i>Demographic Background</i>									
Age	-2.3190***	0.31189	-7.44	-0.0200	0.0122	-1.63	0.4703***	0.1642	2.86
Education									
Special education	-0.5269	16.9582	-0.03	0.2597	0.6692	0.39	9.7036	6.3365	1.53
High school graduate	-24.6122	16.6818	-1.48	2.9657***	0.6654	4.46	41.5547***	6.4365	6.46
Post-secondary/associate degree	-14.8805	18.2926	-0.81	5.1137***	0.7445	6.87	83.4660***	9.0927	9.18
College degree or higher	34.4609	21.8180	1.58	4.6625***	0.8734	5.34	163.9384***	17.9561	9.13
Race									
Native American	-17.4687	49.6692	-0.35	0.1325	1.9237	0.07	5.2426	23.7536	0.22
Asian	57.3725	56.6339	1.01	-3.8100**	1.7111	-2.23	-40.9341**	17.1552	-2.39
White	16.1171*	9.6736	1.67	-0.5949	0.4910	-1.45	-6.2492	5.1003	-1.23
Hispanic or Latino	-26.7465	26.3873	-1.01	0.2442	0.9318	0.26	-10.9634	11.0435	-0.99
Gender (Female)	8.6879	7.7448	1.12	-2.5881***	0.2936	-8.81	-32.0111***	3.7005	-8.65
Veteran	-18.0695	19.4540	-0.93	1.3458*	0.7870	1.71	29.1764**	13.6200	2.14
Primary disability									
Physical impairments	-25.4000	20.0359	-1.27	0.1200	0.8080	0.15	-2.3717	17.4014	-0.14
Mental impairments	-52.5525***	18.0453	-2.91	-2.8383***	0.7448	-3.81	-55.7796***	15.5328	-3.59
Secondary disability									

Sensory/communication impairments	-14.1650	17.5393	-0.81	-3.0087***	0.7232	-4.16	-25.6520***	8.5062	-3.02
Physical impairments	-7.8156	10.0790	-0.78	-1.2690***	0.4236	-3.00	-8.4449	5.6663	-1.49
Mental impairments	-15.1077*	8.2807	-1.82	-0.1644	0.3292	-0.50	-4.0529***	1.5932	-26.29
<b><i>Pre-service Status</i></b>									
Currently employed	7.905325	9.4227	0.84	-0.2275	0.3484	-0.65	11.1346**	5.1178	2.18
Work disincentives	3.0327	2.971615	1.02	-3.8039***	0.1238	-30.47	-41.8839***	1.5932	-26.29
Previous closure/service									
Closed before services	-13.2639	13.1365	-1.01	0.1923	0.5492	0.35	-8.0276	6.8302	-1.18
Closed after services	-22.0510***	8.3032	-2.66	-1.2896***	0.3354	-3.85	-18.4833***	3.8301	-4.83
Participation in Projects with Industry	-41.4231	59.0265	-0.70	-1.4386	2.4348	-0.59	-61.2789**	25.8489	-2.37
<b><i>Employment Services</i></b>									
No. of placement services received	93.0754***	12.3244	7.55	-1.9901***	0.4011	-4.96	-35.1571***	5.7619	-6.10
Constant	292.321***	31.31243	9.34	37.6258***	1.2273	30.66	336.2284***	19.0697	17.63
<i>F</i> -test	15.70***			81.59***			56.09***		
<i>R</i> <sup>2</sup>	.2707			.2601			.2805		

\*significant at .1; \*\*significant at .05; \*\*\*significant at .01; two-tailed tests.

PBC in 2007, Indiana clients experienced higher employment possibilities (odds ratio=1.4991,  $p < .01$ )<sup>11</sup>.

Second, Ordinary Least Squares (OLS) regressions were run to compare three performance indicators of employment outcomes before and after PBC: (1) time to placement, (2) weekly working hours, and (3) weekly earnings (adjusted by inflation). Before regression analyses, a series of regression diagnostics were conducted to ensure the basic assumptions of OLS regression are met. Both White's and Breusch-Pagan tests imply strong concern for heteroscedasticity of the residuals. Thus, robust standard errors were used in regression models. Overall, these three models are significant, with an  $F$  value of 15.7 ( $p < .0001$ ) for the model on time to placement,

---

<sup>11</sup> The interpretation of interaction effects in nonlinear models is still under much econometric discussion (Ai & Norton, 2003; Athey & Imbens, 2006; Greene, 2010; Karaca-Mandic, Norton, & Dowd, 2012). Ai and Norton (2003) argue that in nonlinear models the marginal effect of the interaction term does not represent the magnitude of the interaction effect. The interaction effect depends on all the covariates, and thus requires computing the cross derivative of the expected value of the dependent variable. The statistical significance of the interaction effect should be based on the estimated cross-partial derivative. Puhani (2012) and Karaca-Mandic, Norton, and Dowd (2012) further demonstrate that under difference-in-difference context, the incremental effect of the coefficient of the interaction term could approximate the treatment effect on the treated. We follow this suggestion in this paper.

an  $F$  value of 81.59 ( $p < .0001$ ) for the model on weekly working hours, and an  $F$  value of 56.09 ( $p < .0001$ ) for the model on weekly earnings. All the models also explain substantial portions of the variations of the dependent variables, 27.07%, 26.01%, and 28.05%, respectively.

The regression model on time to placement shows individual employees in Indiana after the use of PBC spent 72 days ( $p < .01$ ) less to achieve employment outcomes, which is consistent with our hypothesis that PBC motivates service contractors to achieve employment outcomes rapidly. The models on employment quality (working hours and wages) demonstrate mixed results. The same as our hypothesis on job retention, individuals in Indiana worked 1.33 hours ( $p < .05$ ) longer than their counterparts weekly during 2004-2006. The hypothesis on wages is partially supported. Weekly wages of Indiana employees increased by \$4.37 after the introduction of PBC, but the difference is not statistically significant even at  $p < .1$  level. However, these small differences on working hours and wages, though meaningful in a statistical sense, are actually of no real policy significance.

## 5.6 Conclusion

The managerial motivation behind all the performance-based management strategies is the phrase that “what gets measured gets done.” The introduction of PBC in human service provision is no exception. By attaching contract compensation to performance achievements, PBC draws contractors to move toward desired service outcomes in a timely manner. Because of the outcome orientation, PBC gives service providers

considerable discretion throughout service process, aiming to encourage innovative and quality customer services that would further result in better service outcomes. This chapter, using a community-outcome perspective, tests these claims by examining the employment service outcomes under two contracting approaches. As predicted, PBC encourages the achievement of employment outcomes in shorter time periods. But the differences between two models in terms of working hours and wages are trivial. It seems “what gets done is what gets measured.” Combining these quantitative evidences together, a conclusion could be made that PBC is better than FFS in that it achieves desired employment outcomes in a more efficient way, without degrading employment quality.

The study in this chapter may suffer from two categories of limitations. First, as mentioned earlier, the local differences between two states were out of full control. In quasi-experiments, there are always risks of comparing different people in different contexts. In propensity score matching, samples are matched on observed covariates, assuming that there are no unobserved differences between the treatment and control groups. This assumption might be too strong to be true in real context –balance on observed covariates may not rule out the role of unobserved differences. To address this potential bias, this research introduced DID on matched samples to further adjust unobserved covariates. Although the paper has some evidence to support the common trend between two states, it still cannot guarantee the similarity in local communities over time where employment outcomes are embedded. Michalopoulos, Bloom, and Hill (2004) remind that comparing groups that are not from the same social context

would potentially bias the estimation of treatment effect, because these groups may be exposed to different local situations and thus unobserved variables. This warning is particularly relevant as we do social program evaluations using comparison groups across service jurisdictions. Heckman et al. (1997) also warns that the bias might be larger in out-of-state comparison groups than in-state comparison groups due to geographic mismatch, such as different geographic location and local labor market.

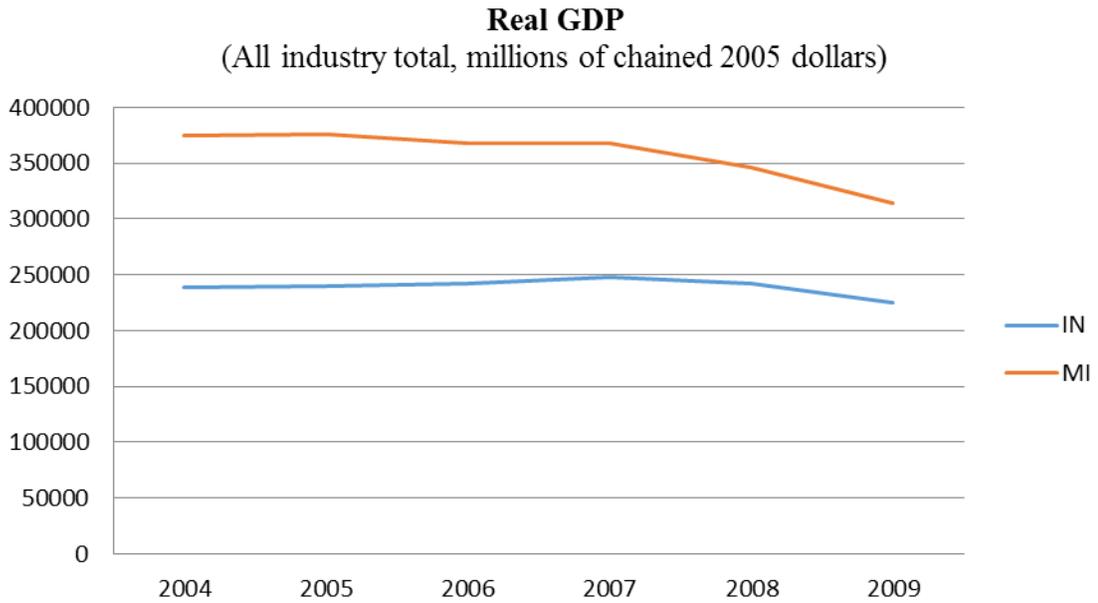
Second, the present research wouldn't compare the differences in several other indicators on employment services, due to the data availability. One major concern on the effectiveness of PBC is the client selection program – contractors may have fiscal incentives to decline severe clients to achieve performance outcomes. Unfortunately, the data used here only records the individuals who had been admitted into service processes. Besides, the costs to achieve employment outcomes under two contracting models were not observed, either. Generally, such costs should include two parts, service costs in the purchase of services from contractors and administrative costs in monitoring contractors. Theoretically, PBC is more economical by shortening service duration and reducing monitoring and reporting. Moreover, the long-term employment effect was not examined, either. In this study here, short-term indicators (working hours and wages at closure) were used as proxies. However, as previously noted, VR services target the long-term stability and welfare of disabled people.

In short, this chapter evaluates PBC effectiveness from a community-outcome perspective. It employs quantitative quasi-experimental methods to compare

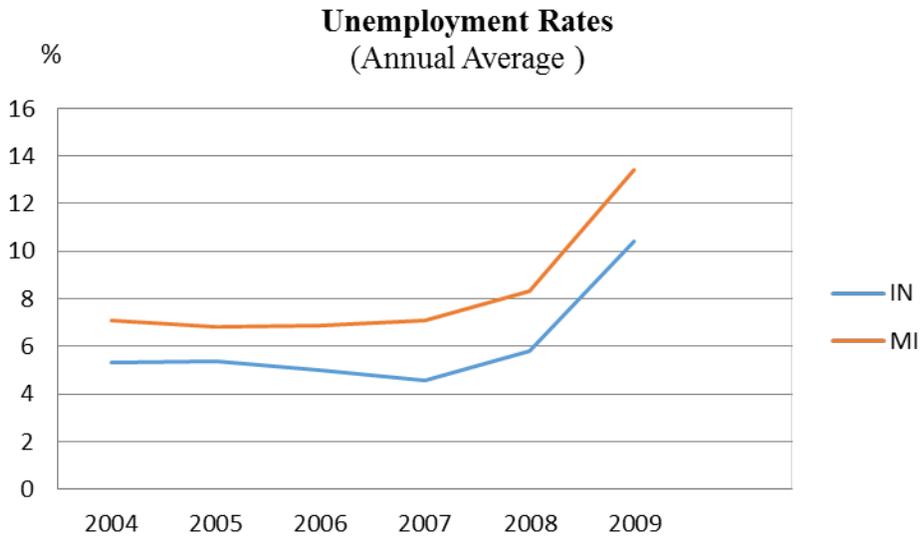
individual employment outcomes under two contracting approaches.

Methodologically, this research design didn't address two things. First, as mentioned above, the quantitative methods used here have a number of limitations, which might bias the robustness of the findings. Second, holding a service-outcome perspective, we might ignore the causes behind the findings here that are directly derived from administrative data and miss rich details of PBC implementation in the real service setting. With these in mind, we turn to next section, qualitatively assessing PBC effectiveness from a participating-organization (i.e., government and contractor) perspective.

Appendix 1. Comparisons between Indiana and Michigan 2004-2009



Source: Bureau of Economic Analysis.



Source: U.S. Bureau of Labor Statistics.

## Chapter 6. The Effectiveness of PBC: Government and Contractor Perspectives

### 6.1 Introduction

This chapter will evaluate PBC effectiveness from a participating-organization perspective. As mentioned in the previous chapter, network effectiveness should also consider the organizations involved in the contractual networks. Indeed, their survival and success affect the effectiveness and sustainability of a network. Often, individual organizations have different interests and hold different motivations in participating in service-provision networks. How the network structure (i.e., PBC arrangement) defines and influences the behaviors and incentives of participating organizations is the central question of this chapter. Generally, two types of organizations are involved in the provision of VR services, government and service contractors. Thus, the chapter will assess the organization-level effectiveness of PBC from these two perspectives. Particularly, the chapter holds a street-level lens and uses qualitative methods in the examination of PBC implementation by the two actors.

### 6.2 Street-level Perspective in Policy Analysis

Policy implementation is complicated. Policy implementation, as Bardach (1977) defines, is “(1) a process of assembling the elements required to produce a particular programmatic outcome, and (2) the playing out of a number of loosely interrelated games whereby these elements are withheld from or delivered to the program assembly on particular terms” (57-58). Traditional public administration literature has

shown “the complexity of joint action” (Pressman & Wildavsky, 1973) and the “great difficulty of organizing cooperative activity on a large scale” (Derthick, 1972) in converting policy intents into policy actions.

Scholarly research on policy implementation has identified two analytical approaches: top-down and bottom-up<sup>12</sup>. The top-down approach, or what Elmore (1979) calls “forward mapping,” begins at the top of the process, with a clear emphasis on policy designers as the central actors. It traces policy implementation along the hierarchical structure within traditional bureaucratic system and explores the ways to guide and constrain the behavior of civil servants and target groups in realizing the policymaker's intent. In this vein, Mazmanian and Sabatier (1983) consider policy implementation as “the carrying out of a basic policy decision, usually incorporated in a statute but which can also take the form of important executive orders or court decisions ... ” (20). Thus, the prescription developed by this school of thought to ensure faithful implementation generally centers on formal organizational structures, authority relationships, control and regulations, administrative responsibility, and so on. The implicit assumption behind this approach is that policymakers control the administrative and political resources that are needed to guarantee policy implementation.

---

<sup>12</sup> There is another so-called “third-generation” approach of policy implementation study, claiming for a synthesis of both top-down and bottom-up frameworks (e.g., Matland, 1995; Sabatier, 1986). But there is actually little advancement in this regard. See O’Toole (2000) for more details.

On the contrary, the bottom-up approach, or “backward mapping” (Elmore, 1979), emphasizes policy implementation at local level and considers the role of local actors in the interpretation of grand policy goals. Due to local differences, the remote control enforced by policy makers is inevitably incomplete and thus lower-level administrators would enjoy certain discretion in turning policy intent into policy action. Based on this, this school of thought argues that it is how lower-level administrators use their discretion to adjust to local context that determines the real meaning of a policy. This strand of research typically explores the dynamics on the recipient level and analyzes the real causes that influence the mutual adaptation of a policy to its local organizational setting.

“The crucial difference of perspective”, as Elmore (1979) writes, “stems from whether one chooses to rely primarily on formal devices of command and control that centralize authority or on informal devices of delegation and discretion that disperse authority” (605). The top-down approach is concerned more with compliance, while bottom-up approach values bargaining and compromise. In social policy, Berman (1978) distinguishes “federal macro-implementation” from “local micro-implementation,” tailored to the dichotomous institutional nature of policy implementation – the federal determines the grand picture, while the local organizations adapt the mission to local setting and deliver the concrete services. In this way, “the net result is that the effective power to determine a policy’s outcome rests with local deliverers, not with federal administrators” (157). Based on the elaboration on control strategy in human service contacting in chapter two and the

quantitative findings in chapter four, it is not surprising that the chapter here tends to rely on the “backward mapping” logic and the street-level perspective in particular.

Lipsky (1980) defines these local public service organizations (such as schools, hospitals, police offices and welfare agencies) as street-level bureaucracies and the front-line workers within as street-level bureaucrats (SLBs). He argues that working in the conditions of huge caseloads, ambiguous agency goals, and inadequate resources, SLBs’ exercise of discretion is inevitable in the day-to-day implementation of public programs. When combined with substantial discretionary judgment and the requirement to interpret policy on a case-by-case basis, the gap between policy intent and policy action can be substantial. Therefore, “the decisions of street-level bureaucrats, the routines they establish, and the devices they invent to cope with uncertainties and work pressures, effectively become the public policies they carry out” (Lipsky, 2010, xiii).

The role of SLBs can also be understood from an organizational behavior perspective. Locating at the boundary between public welfare agency and its external constituency, SLBs actually play a critical boundary-spanning role in information processing and external representation (Thomson, 1967; Prottas, 1978). “Information from external sources comes into an organization through boundary role, and boundary role link organizational structure to environmental elements, whether by buffering, moderating, or influencing the environment” (Aldrich & Herker, 1977, 218). In the information processing function, boundary roles gather, transmit, and

interpret information from external environment for internal organizational components. In the external representation function, boundary roles serve resource acquisition and institutional legitimation.

Summarily, SLBs typically perform these two functions, acting as a mediator between welfare agencies and targeted service receipts and resulting in bureaucracy's dependence on its SLBs in people processing. Such dependence constitutes a source of power. "To the extent that information access and control is a power resource, boundary spanners are in an excellent structural position to convert this resource into actual power. ... Their power is further enhanced to the degree that the nature of the task assigned the boundary role makes routinization of the role difficult, if not impossible" (Aldrich & Herker, 1977, 227).

The massive use of contracting in social service delivery complicates the original connotation of SLBs. Under the contracting regime, contractors have partially taken over the functions previously performed directly by social workers. Now, contractors directly work with clients and provide services to achieve goals set by government agencies, while social workers turn to determine the client eligibility and monitor contractor behavior. Collectively, these two actors become new SLBs (Smith & Lipsky, 1993). This notion of new street-level bureaucrats is consistent with Hjern and Porter's (1981) suggestion of using "implementation structures" as a unit of analysis for studying "purposive action within a framework where parts of many

public and private organizations cooperate in the implementation of a programme” (214).

Lipsky’s construct of SLB provides a useful perspective in studying social policy implementation. According to Brodtkin (2003), “[t]his approach is most valuable when policy implementation involves change in organizational practice, discretion by frontline workers, and complex decisionmaking in a context of formal policy ambiguity and uncertainty” (151). Using the street-level approach, we could explore how street-level actors put policy into practice, more specifically, “what street-level organizations construct as policy through their informal practices, how they do it, and why they produce policy in the ways that they do” (Brodtkin, 2011, 199). The robustness of this approach has been shown by numerous studies of welfare reform in a variety of policy areas (e.g., Brodtkin, 2011; Keiser, 2010; Meyers, Glaser, & Donald, 1998; Riccucci, 2005).

This study relies on this street-level perspective and explores the mutual adaptation of PBC to its local organizational setting. How is performance-based contracting being implemented? Do PBC and the incentive system associated work as intended to motivate SLBs to improve service outcomes? Particularly, this study would examine: How do SLBs respond to PBC? How do they use their discretions under PBC? Does PBC motivate them to work in the “right” direction, both short-term and long-term? Do they use discretions to gaming the PBC? If so, how negative are they and how do public managers deal with this?

### 6.3 Vocational Rehabilitation Context and Data Collection

The context for examining the question of how PBC is implemented is state vocational rehabilitation programs. As described in chapter three, in rehabilitation services, VR counselors in the state agencies work with clients on eligibility determination and progress monitoring, while employment specialists at service contractors coach and counsel VR clients to find jobs. Thus, The SLBs or front-line workers in this setting are counselors and employment specialists.

The main research method here is semi-structured interviews (interview questions attached in Appendix 2). Qualitative data for the research was collected with staff from local VR offices and service contractors in Indiana. Certainly, the ability to gain access to local offices and contractors by the researcher was indeed a criterion in sample selection. Generally, VR local employees were invited to have 30-40 minute talks concerning their experience on PBC from their own perspectives. Interviews were operated in spring and summer 2013, mostly through telephone, except for one face-to-face interview with a counselor conducted in a local VR office in Indianapolis. Detailed notes were taken based on interviewees' inputs.

Table 18 describes the distribution of interview samples. It is important to note that the goal of the interview is not to generate statistics about a population. Thus, it did not follow a strict probability sampling. Rather, it uses a snow-ball sampling strategy.

Before interviewing SLBs, interviews with state VR program managers and area supervisors were conducted to collect background information and “set the stage” for later research. Also, content analysis of key documents relating to service contracts, annual state program report, and meeting minutes were conducted. Data from all these sources are used jointly for analysis.

Table 18. Distribution of Interview Samples

	Indiana
VR counselors	5
Program managers (contractors)	4
Employment specialists (contractors)	3

#### 6.4 Findings from VR Agency Perspective

Views from government side on PBC effectiveness are mostly positive. Counselors were impressed with high costs and poor placement outcomes under the traditional FFS arrangement. Under FFS, due to the mechanism of reimbursement for services provided rather than service outcomes, there were frequently large amount of job development services, sometimes more than necessary, by vendors before client placement. So counselors saw vendors keeping providing services to clients, but without placing them. FFS also means labor intensive for counselors: they had to work closely with clients and vendors. Counselors needed to meet with clients very frequently (e.g., every month) to confirm their situations. Vendors were asked to report to counselors intensively (every 30 minute) when providing detailed services to

clients. There were also a number of administrative procedures for vendors to go through, such as monthly billing and reporting. In contrast, PBC is easier to manage from government perspective. Under PBC, counselors enjoy more flexibility and less job burden. They stay involved, but do not track vendor behaviors that closely. Counselors only need to authorize milestones at the beginning and then verify clients' achievement, rather than regulate detailed services vendors provided.

When talking about clients' employment outcomes, counselors were mostly impressed with PBC's effect on better employment outcomes with less costs. Because of the financial incentives related to the milestone payment, PBC directs contractors to better employment outcomes. As mentioned earlier, VR services roughly include three steps: job development, job placement, and job retention. PBC is very effective in promoting placement and keeping the job (at least to case closure). For example, in Indiana milestone system, to receive the payment on job retention milestone, vendors have to help regular clients keep their jobs for at least 90 days. Indeed, success in longer-term placements has been a persistent challenge in VR programs. In this way, increased placement possibility and job tenure are most evident effectiveness of PBC. However, the changes on employment wages, benefits, job match, and other aspects of life quality under PBC are not emphasized by counselors. These are those vocational characteristics that are not specified explicitly by the PBC milestones. In these areas, both FFS and PBC might perform quite equally and lack impressive differences. One explanation might be that vendors under either contracting arrangement have the same non-financial incentives to perform these uncaptured

performance areas. In addition, counselors have no evidence to track clients' long-term employment and stability.

Also, PBC pushes vendors to move across the service process to hit the milestones and receive payment. Because services would no longer be directly reimbursed by funding agencies under PBC, vendors have no financial incentives to hold clients and keep providing services to them, but pursue rapid job search and accelerated placements. Given that, counselors all agreed that PBC greatly reduces service waiting time and time clients spent to achieve placement.

Up to the VR program level, PBC improves program costs in two ways. On one hand, due to "paying for performance" of PBC, only desired service milestones would be paid by funding agencies. In this way, funding agencies actually shift financial risks to vendors. Therefore, there is substantial saving on service costs. On the other, funding agencies no longer need to track vendors very closely to monitor their behaviors in the service process. Less administrative monitoring means a decrease in administrative costs, which is usually one third of total program costs. With these savings, VR agencies, which often suffers from overloaded service applicants, would extend their program capacity and serve more clients.

There are certainly a number of managerial concerns/challenges some counselors saw in the implementation of PBC. When using PBC, government has less control of the service process. Unlike FFS, PBC grants huge amount of discretion to vendors and

substantially reduces administrative control in the service process. Under PBC, counselors only need initial authorization of service milestones and verification of milestone achievement, being largely out of touch with clients and the service process. Thus, PBC changes the role of counselor from director of services to a payer of services. It also shifts the responsibility of vocational assessment and planning from counselors to contractors. Some interviewees felt that PBC discounted the importance of the counselor role and government involvement.

Such diluted government control, if not well managed, might lead to problems. The essence of PBC, according to Wedel and Colston (1988), is to use specified rewards for meeting or exceeding contract objectives and penalties for failure to meet them. This actually points to the critical role of incentive structure design. However, building consensus on outcome/milestone measurement and payment rate within funding agencies and with diverse contractors is not an easy work. For example, the milestones currently being used in contracted employment services are mostly easy-to-observe performance measures, such as one-week placement, 90-day employment. Counselors did admit there should have some other milestones to consider. Some also mentioned whether to include some invisible measures related to positive life quality changes, in order to fight against potential creaming by contractors. Counselors observed that contractors were better at meeting some milestones than others. Also, in the long run, how to adjust milestones and payment rate would be challenging. As such, there is a risk that the incentive structure associated with PBC, if not appropriately designed, might incur undesired vendor behaviors.

With regards to creaming/gaming, counselors did put forward concerns for this negative potential of PBC. However, most counselors believed creaming did not occur frequently. They didn't see strong evidence of creaming or less outcomes, although these observations were directly from intuition, rather than systematic data analysis. Part of the reason might be that the two-track payment system (paying vendors with a higher rate when serving people with higher degree of disabilities) effectively address the creaming problem. However, counselors noticed vendors supported PBC generally due to the flexibility vendor have in the service process and complained the financial risks vendors face. This financial pressure had make some vendors drop off due to poor service performance. So, this pressure might make vendors less engaged in service process, or even lead to deterioration.

Generally, VR counselors were consistently satisfied with PBC. However, they did admit that PBC might not work in every situation. Hourly payment is still needed, particularly when considerable specific services are required. Thus, even though system-based conversion from FFS to PBC is necessary, some usage of hourly payment should still be allowed.

#### 6.5 Findings from Contractor Perspective

Unlike VR counselors, contractors hold a quite mixed attitude towards PBC. The same as counselors, contractors welcome the flexibility associated with PBC in serving clients. Under FFS, contactors were closely monitored by counselors and

funding agencies. When working under PBC, they could take activities they think are necessary to achieve milestone outcomes. Because only milestone outcomes would be evaluated by VR agencies, vendor behaviors are free from strict administrative control. All contractors interviewed were very much impressed with the decrease in the amount of time they spent in administrative reporting and paperwork. In general, contractors support PBC in that it promises more rapid initial assessments, faster VR authorizations (due to less administrative paperwork), less paperwork and reporting, and less scrutiny by counselors. Although more discretion does not mean better service quality or more innovative service methods, contractors are more likely to serve clients in the ways that they believe are best practices.

PBC also encourages contractors to work with clients more intensively, at least in the early placement process. Contractors reported that they conducted more in-person contacts and spent more evaluation time with clients under PBC, in order to place clients in good jobs that clients would want to keep. However, such improvements in service quality with clients might be due to the motivation of achieving milestones and securing funding in the shortest time. The incentives of rapid placement might lead to temporary jobs to gain reimbursement at earlier stages at the expense of longer-term jobs. Some clients left jobs because they did not like them or failed to accommodate to workplace. It partly indicates that the employment support services contractors provide in the service process before placements might not be sufficient. In short, PBC might detract from the desired outcomes to the extent that contractors feel pressured to place clients in jobs sooner at the expense of job fit. Also,

contractors might use enhanced interactions with clients early in the process to terminate unmotivated or highly difficult client earlier in the process to avoid potential risks.

Contractors were very concerned with the financial risks shifted from funding agencies: contractors would not be reimbursed until milestones are achieved. Some contractors complained that PBC undermined the vocational philosophy, changing from serving clients' overall vocational needs to a narrow focus on meeting milestones. Although they didn't admit they selected clients based on the possibility of future employment, they did emphasize their employment programs had budgetary constraints. Due to the high financial pressures under PBC, contractors have to consider the cost and risk of serving a certain client. As for employment outcomes, contractors emphasized that because of financial reasons, they would pay attention to clients' milestone achievement that would lead to payment. However, they didn't think PBC would induce them to engage in creaming/gaming behaviors, achieving measured performance at the cost of unmeasured ones. Contractors believed other non-financial incentives under FFS and PBC were of no difference. Thus, they didn't notice significant different in clients' employment quality, such as wage, job match.

Generally, contractors understood the rationale behind the change from FFS to PBC. However, contractors were anxious about the high financial risk PBC would pose on their running employment programs, despite the flexibility they enjoyed under PBC. Weighing these two things, contractors still preferred FFS in providing employment

services. At least, some contractors proposed a hybrid contracting model with a combination of FFS and PBC. Funding agencies should use FFS for difficult clients, while PBC for regular ones.

## 6.6 Conclusion

Here, let's summarize the findings from this qualitative piece. For VR counselors and service contractors, the first impression they have concerning PBC is the flexibility, in either administrative or service sense, PBC brings. On the counselor side, they no longer need to monitor contractors very closely throughout the service process. Rather, they just authorize milestones and verify whether they have been really achieved by clients, becoming a little away from clients and vendors. On the contractor side, without such administrative control, they could serve clients in the way they believe to be the most professional one. Thus, in this regard, both parties expressed strong satisfaction with PBC.

Second, PBC, through financial incentive restructure, directs contractors away from service delivery per se towards milestones along job development, job placement, and job retention. Following this logic, contractors, in order to receive service reimbursement, devote much attention to these milestones. So counselors could observe clients under PBC find jobs sooner and are more likely to keep jobs at least till reaching the milestone on case closure. But PBC's change on other areas of employment outcomes (wage, job match, etc.) are not impressive to counselors.

Given that PBC has no touch on these indicators, contractors explained the incentives in these areas might be equal in FFS and PBC.

Third, there are some unsupported clues on creaming/gaming. For contractors, the financial incentives of rapid movement across milestones might lead to clients' temporary jobs at earlier stage at the cost of job fit and longer-term jobs. However, whether and to what extent would PBC differ from FFS in this regard are ambiguous. For counselors, although there are concerns for creaming/gaming, they actually have no evidence to show these strategic behaviors indeed exist and play roles.

At the beginning of the last chapter, this project aims to evaluate PBC effectiveness from two perspectives using different methods. The previous chapter employs a quantitative quasi-experimental method to compare the impacts of two contracting approaches on individual employment outcomes. The present chapter, through qualitative semi-interviews, explores how PBC was implemented by street-level actors, both VR counselors and contractors. How do these two perspectives and methods jointly help us understand PBC effectiveness? This is the topic of the next chapter.

## Appendix 2. Interview Questions on the Effectiveness of Indiana RBF

### *Questions for VR Counselors:*

1. What are the major differences between FFS and RBF, from your perspective?
2. What do you like most about RBF and what do you like least?
3. How might RBF change service activities and process?
  - (1) Service method (best practices: more job development, search, and match)
  - (2) Client selection (two-tier payment useful?)
  - (3) Service costs (purchase costs and administrative costs)
4. How might RBF change clients' employment outcomes?
  - (1) Employment results
  - (2) Time to placement
  - (3) Employment quality (wage, working hours, job match, customer satisfaction)
  - (4) Long-term employment and stability
5. How might RBF change your work as a counselor, and relationship with service providers?
  - (1) Work with clients
  - (2) Oversight service providers
6. How do service providers view RBF, from your perspective? What are their concerns?
7. All things considered, which funding mechanism does you actually prefer, RBF or FFS?
8. What might be the potential room to improve RBF?

### **Questions for Contractors:**

1. What are the major differences between FFS and RBF, from your perspective?
  - (1) RBF: Like most? Like least?
  - (2) FFS: Like most? Like least?
2. How might RBF change service process?
  - (1) Client selection
  - (2) Best practices
3. How might RBF change clients' employment outcomes?
  - (1) Possibility of employment
  - (2) Time to placement
  - (3) Employment quality (wage, working hours, customer satisfaction)
  - (4) Long-term employment and stability
4. How might RBF change your work as a service provider, and relationship with counselors?
  - (1) Work with clients (consumer evaluation, job development, job search, documentation, and in-person contacts)
  - (2) Oversight from VR agency
5. All things considered, which funding mechanism does you actually prefer, RBF or FFS?
6. What might be the potential room to improve RBF?

## Chapter 7. Two Faces of Contracting, Two Kinds of Control

### 7.1 The Effectiveness of PBC as a Formal Arrangement

No doubt, PBC is mostly a formal contracting endeavor. Through restructuring formal contract design, or more precisely, changing the financial incentives from “paying for process” to “paying for result,” PBC draws contractors’ attention towards the results of service delivery, rather than service delivery per se. However, as discussed in chapter three, the appropriate contract design for human services is very daunting. Human services always feature their ambiguousness in both task programmability and outcome measurability. This puts human services mostly within what Weisbrod (1988) calls Type II dimension of service attributes<sup>13</sup>. In this way, neither behavior-oriented contracts (FFS) nor outcome-oriented contracts (PBC) fit seamlessly with human services. Given this, the question becomes which contract arrangement is less risky.

The entire research is to evaluate the effectiveness of PBC in human services, using Indiana vocational rehabilitation program as a case. Towards this goal, the previous two chapters employ two perspectives and two research methods. In chapter five, PBC effectiveness was assessed from a service outcome perspective, based on a

---

<sup>13</sup> Weisbrod (1988) differentiates between Type I dimension of service attributes (that are relatively easy to monitor or assess) and Type II dimension of service attributes (that are relatively difficult to monitor).

quantitative quasi-experiment to compare the impacts of PBC and FFS on individual employment outcomes. In chapter six, PBC effectiveness was explored using qualitative semi-interviews with VR counselors and contractors about PBC implementation. Clearly, both quantitative and qualitative methods have weaknesses. Quantitative and qualitative methods follow different paradigms, assumptions and have different strengths in information processing (Firestone, 1987). Quantitative methods, based on a positivist paradigm, “produce factual, reliable outcome data that are usually generalizable to some larger population.” In contrast, qualitative methods, grounded on a phenomenological paradigm, “generate rich, detailed, valid process data that usually leave the study participants’ perspectives intact” (Steckler, McLeroy, Goodman, & Bird, 1992, 2). Accordingly, a research strategy advocated in methodology literature is triangulation, “the combination of methodologies in the study of the same phenomenon” (Denzin, 1978, 291). It has been suggested to see quantitative and qualitative methods as complementary, used to compensate for the limitations of each other and cross-validate to gain greater accuracy and confidence in judgments (Jick, 1979; Mathison, 1988).

The present research tends to follow this “triangulation” logic, through collecting and analyzing different kinds of data bearing on the use of PBC in VR employment services. Not surprisingly, the findings from the two methods, through illuminating, are not exactly the same, but some general conclusions could be derived. The effectiveness of PBC, from a service outcome perspective, means clients find jobs that are permanent, pay good (above minimum) wage, and match their interests. From

an organizational perspective, effectiveness might be close to administrative efficiency and flexibility in running VR programs and providing employment services.

In the service effectiveness sense, PBC performs better in the areas that are captured by milestones. In other words, compared with FFS, PBC is more likely to achieve the milestones. Because of the financial incentives, contractors push clients they serve to move across the milestones rapidly. Although PBC has little impact on unmeasured areas, there is no strong evidence of creaming/gaming. Contractors might be involved in strategic behaviors in some cases, but those behaviors haven't been found to result in deterioration in service outcomes.

In the organizational effectiveness sense, PBC are well endorsed by funding agencies for its efficiency and flexibility. VR counselors become relatively free from intensive work with clients and contractors and enjoy much flexibility in managing the service process. Funding agencies spend the money and get the results they want, without seeing severe unintended outcomes. For contractors, PBC would be a double-edged sword. They support PBC in that it allows more exercises of professional discretion, but complain high financial risks they burden. This risk is indeed a big managerial challenge. If not appropriately handled by public managers, it might force contractors to engage in more strategic behaviors.

To an extent, PBC seems more promising in both service effectiveness and organizational effectiveness. However, it also implies PBC effectiveness might not be well-rounded and should not be exaggerated. After reviewing the use of PBC in federal agencies, GAO (2002) doubts “whether agencies have a good understanding of performance-based contracting and how to take full advantage of it” (2). Accordingly, a policy question would arise: how to improve the effectiveness of PBC, or how to take advantage of it?

From the formal contracting perspective, the most direct response would be optimizing the PBC design, such as fixing performance measures, redefining the connection between performance indicators and contact compensation, and changing incentive structures. For example, Hill’s (2006) study of casework task configurations in welfare-to-work programs finds that the separation of measurable and unmeasurable tasks among frontline workers would contribute to program effectiveness. Heinrich and Choi (2007) suggest changing performance measures periodically before contractors learn the ways to gaming the measures. This would cause a “competition of learning.” When launching a PBC system, both government and contractor start their learning activities, the pros and cons of that system. If government learns faster, they could find ways to fix the problems. If contractors learn faster, they might gaming the system. Anyway, both suggestions warn that PBC should be used very carefully.

However, these technical efforts on restructuring PBC systems would hardly be free themselves from the puzzle of introducing performance management to human service contracting mentioned previously. More broadly, this illustrates what Van Thiel and Leeuw (2002) call “performance paradox” in the public sector – “characteristics of the public sector can be counterproductive to developing and using performance indicators” (267). In this way, the use of PBC in public human service programs might always be at the risk of “rewarding A, while hoping for B.” A more feasible way to optimize PBC endeavor might be a relational one.

## 7.2 Managerial Implications from a Relational Contracting Perspective

The theoretical framework in chapter three presents two faces of contracting, formal and relational. Relational contracting perspective highlights the role of relational sanction and social interaction in contractual fulfillment. It suggests relying on relational exchange as a social control mechanism in contracting management. These two faces of contracting remind that the coexistence of these two mechanisms that public managers should devote themselves to simultaneously. To some extent, such combination of formal and informal contracting reflects the nature of contracting management in public administration context: well-planned and written contracts to meet the formal accountability demand, and negotiation and discretion to satisfy the flexibility concerns in service delivery (DeHoog, 1990).

However, the efforts on PBC innovation and implementation tend to ignore the relational contracting side. As we’ve seen, the formal effort of using PBC in

vocational rehabilitation services was disturbed by the highly uncertain nature of employment services. Therefore, we could see incomplete improvement in employment performance, mostly in targeted performance areas, and the risk of contractor opportunistic behaviors. Rather than the attempt to use other formal devices to awkwardly improve PBC effort, this project suggests introducing relational contracting, with a focus on relationship and trust building, as a supplement.

Relational contracting, as Sclar (2000) suggests, “transform[s] the notion of contracting from a market-based arrangement to one rooted in interorganizational trust” (123). This notion is also termed by sociologists as “embeddedness,” to recognize the role of socially embedded relationships in economic exchange (e.g., Powell 1990; Uzzi, 1997). Granovetter (1985) argues that formal exchanges would “become overlaid with social content that carries strong expectations of trust and abstention from opportunism” (490). Exchanges characterized by trust are generally found to be more successful (Dyer, 1997; Klein Woolthuis et al., 2005; Ring and Van de Ven 1994).

Therefore, relational exchanges, based on social components (Macneil, 1980), are always associated with a higher level of trust. Here the research adopts Rousseau, Sitkin, Burt, and Camerer’s (1998) definition of trust: “a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behavior of another” (395). They also identify two preconditions for trust to arise: risk (or uncertainty) and interdependence. Risks in exchanges create

opportunities for trust. Trust would not be needed if exchanges could be conducted with complete certainty. Although high risks might force two parties to seek other alternatives, interdependence between parties would glue them together.

Interdependence means that the goals of one party's could not be achieved without the participation of the other's. These two conditions further imply the relevance of the discussion on trust here to service contracting. In human services, governments heavily rely on third-party actors to deliver various services to citizens. However, due to the uncertain nature of human services mentioned above, contracting performance is at the risk of contractor misconducts.

The role of trust in interorganizational exchanges and collaborations has been discussed extensively by scholars from a variety of disciplines such as sociology, psychology, and economics. From a sociological perspective, trust acts as a functional alternative to rational prediction for the reduction of complexity in social life (Lewis & Weigert, 1985; Luhmann, 1979). From a transaction cost viewpoint, trust reduces transaction costs by reducing both ex ante and ex post opportunism (Williamson, 1993). Ostrom (1998) suggests trust and reputation for trustworthiness as core factors in collective actions, potentially reducing uncertainty and transaction costs.

Management scholars McEvily, Perrone, and Zaheer (2003) propose trust as an organizing principle, structuring and mobilizing organizational components. In the structuring role, trust affects "the development, maintenance, and modification of a system of relative positions and links among actors situated in a social space" (94). In the mobilizing sense, trust "involves motivating actors to contribute their resources,

to combine, coordinate, and use them in joint activities, and to direct them toward the achievement of organizational goals” (97). In short, as Zand (1972) summarizes, trust “conveys appropriate information, permits mutuality of influence, encourages self-control, and avoids abuse of the vulnerability of others” (238).

The efforts on trust conceptualization tend to acknowledge its multi-faceted nature (Williamson, 1993). Lewis and Weigert (1985) distinguish three dimensions of trust: cognitive, emotional, and behavioral dimensions. To them, cognitive familiarity, emotional bond, and behavioral enactment construct the sociology base of trust. Zucker (1986) also identifies three modes of trust production: (1) characteristic-based, (2) process-based, and (3) institutional-based. Characteristic-based trust can be formed on the basis of individual social characteristics such as ethnicity and background. Exchange partners with similar characteristics are easier to engage in collective actions in that they might believe such exchange would satisfy both parties. Trust can also result from previous and expected future exchanges, i.e., a record of reputation. In institutional-based trust, exchanges are embedded in social practices and trust is thus tied to broad societal institutions.

This paper builds on Zucker’s classification. Indeed, characteristic-based trust has been well observed. For example, as mentioned in the previous chapters, in human service contracting, public managers tend to trust nonprofits’ social-mission driven would prevent nonprofit contractors’ opportunistic behaviors. This mission/value alignment produces characteristic-based trust, which makes public agencies incline to

partner with nonprofit contractors. The following paragraphs focus more on process-based and institutional-based trust and discuss implications that public managers might consider when optimizing PBC efforts from a relational contracting perspective. A summary is provided in Table 19.

Table 19. Mode of Trust Production and Implications for PBC

<b>Mode of trust production</b>	<b>Basis</b>	<b>Implications for PBC efforts</b>
Characteristic-based	Individual attributes	<ul style="list-style-type: none"> <li>• Contractor’s nonprofit status</li> </ul>
Process-based	Past or expected exchanges	<ul style="list-style-type: none"> <li>• Collaboration and negotiation</li> <li>• Time boundlessness</li> </ul>
Institutional-based	Social structures	<ul style="list-style-type: none"> <li>• Professionalism</li> <li>• Best practice</li> </ul>

Source: Zucker (1986).

*Collaboration and Negotiation*

Reciprocal obligation should be a key principle in the use of PBC. Due to the complicate and dynamic nature of PBC, the implementation of PBC as a system-based change would not succeed without the commitment from all stakeholders. After all, the central goal of PBC is to meet client needs while addressing the financial realities of both funding agencies and service providers. For example, as found in the last chapter, the high financial risks burdened by service contractors under PBC, if not well moderated, might lead to gaming or other strategic behaviors. One way to

address the dysfunctional response is to collaborate with stakeholders throughout contracting process. Such collaboration itself acts as a sign of commitment and a tangible expression of mutual trust.

The collaboration should start in contract planning and design stage. It requires the participation from three groups of stakeholders: funding agencies, service contractors, and clients (O'Brien & Revell, 2005). Because the process and outcome of human service delivery are relatively uncertain, stakeholders should reach at least some consensus on incentives and disincentives associated with the PBC design, such as essential milestones, fee structure. Also, the formal contract can be seen as a coordination mechanism to specify what goals all parties aim for and how they want to achieve these goals. The emphasis here can be more on the positive (shared mission, goals, etc.) than the negative (legally enforceable provisions and penalty). Overall, this collaborative planning should ensure that the design addresses each party's concerns and eliminate possible resistance to change. It creates a transparent and participatory process that could enhance the feasibility of PBC design and the likelihood of full implementation.

The development of PBC is also a learning process, for both funding agencies and contractors. Throughout PBC implementation, tensions and conflicts could be anticipated. Thus, ongoing negotiation and system modifications are necessary. At the early conversion (from FFS to PBC) stage, substantial time and resources are needed by the funding agencies to provide technical assistance and training for contractors

and develop shared commitment with them. After the shared development phase, the use and availability of multiple communication strategies to disseminate information would enhance implementation. The stakeholders still need to meet periodically to assess the implantation and make recommended changes. Funding agencies might hold annual program feedback meetings or annual on-site visits to collect contractors' and clients' inputs. These regular and stable interactions reduce opportunistic behaviors and support the development of commitment and adaptation. In this way, trust is formed incrementally and enhanced through repeated interactions. (Dyer, 1997; Gulati, 1995; Lee et al., 2012; Ring & Van de Ven, 1994).

#### *Time Boundlessness*

To an extent, contractor behaviors reflect their expectations for future exchanges. In discrete time-bound transactions, people might respond to calculations of short-term advantage. In contrast, open-ended contracts imply potential benefits from future collaborations and thus provide a safeguard against opportunistic behaviors. Open-ended contracts not only convey a sign of commitment and mutual trust at the beginning of exchanges, but promote trust formation and enhancement in the long run due to repeated interactions mentioned above. Researches find exchanges that operate for a pre-specified duration would behave differently from those under a setting of continuing relationships and interdependence. Axelrod (1984) suggests compared with open-ended contracts, time-bound contracts are less likely to be self-enforcing due to the lack of a "shadow of the future." Reuer and Arino (2007) confirm that

time-bound contracts would cause a greater threat of opportunistic behavior and contribute to contracting complexity.

Taken together, the arguments here suggest the use of open-ended contracts or at least longer-term contracts. Currently, in vocational rehabilitation programs studied in the present project, for example, service contract duration is usually one year, with an option of one-year extension contingent on satisfactory performance. Under this specified short duration, contractors might consider short-term opportunistic behaviors. Of course, continuing relationships do not necessarily mean nice. Relational sanctions would not always produce cooperation. Rather, they might lock funding agencies into dependent positions (Williams, 1983). However, this point here does not exclude other formal enforcement mechanisms such as performance assessment and financial auditing, but suggest the combination of open-ended contracts with other formal control tools.

### *Professionalism*

Professionalism achieves social legitimacy through specialized expertise and qualifications. Under information asymmetry, professionalism acts as a signal of quality, ensuring that professionalized organizations are in compliance with established social expectations and professional standards. It is termed by Ouchi (1979) as a “ritualized, ceremonial forms of control” (844). Professionalism means only a selected organizations and individuals who have gone through

professionalization processes could be allowed to participate in the service program operation and service delivery process. For human service organizations, accreditation is a kind of quality assurance that an organization meets the quality standards established by the profession. Accredited organizations are required to follow similar service procedures and occupational norms, which convey an assurance of quality and credibility. For service workers, professional schooling and membership would be a channel to internalize the desired attitudes, values, and beliefs. For nonprofit human service organizations, professional values are also reflected by a professional workforce that an organization uses in service jurisdictions and management. These “organizational professionals” (DiMaggio & Powell, 1983) generally hold occupational norms and standards, implying a higher degrees of professionalism in organizational operation.

#### *Documentation of Best Practices*

Another somewhat relevant to professionalism is the “best practice” approach. Funding agencies might document and disseminate periodical reports, identifying the contractors with best service outcomes and their best practices in service delivery. This would create some informal pressure on service contractors with poor performance in their profession. However, the approach builds on the assumption that service contractors have relatively strong self-motivation to provide better services and care about professional recognition (Else et al., 1992). If so, they would wish to learn from leading organizations and improve their own performance.

### 7.3 Conclusion: Control, Trust, and Contracting Management

In the United States, government contracting is widely and durably used as an indirect government tool in the landscape of service delivery and policy implementation. This governing by contracting model has fundamentally redefined the U.S. governance system, in both political and managerial senses. It also highlights the imperative of contracting management to ensure high-quality results. However, public managers are often frustrated by their insufficient management capacity while working with contractors. To address this “smart-buyer” challenge, public management scholarship and practice in past three decades have conducted a huge amount of exploration of effective contracting management. Inspired by performance management movement, PBC represents one of the most recent efforts. PBC incorporates performance measures in contract specification and makes contract compensations attached to contractors’ performance achievement. Theoretically, PBC promises quality services, better outcomes, and less monitoring.

Given the potential benefits, governments at all levels have shown substantial and continuous enthusiasm for PBC. In human services, particularly, state and local governments have expressed growing interests in using PBC in their service acquisition. However, the burgeoning popularity of PBC lacks sufficient evidence to show its promised benefits are actually achievable. In particular, the introduction of PBC into human service systems needs to address the effectiveness problem (whether PBC produces better results) and the capacity problem (how to use PBC and lead

interorganizational change). The present research mostly focuses on the first problem, while the findings here might shed some light on the second problem.

After building the theoretical framework which incorporates the literature on formal and relational contracting, this research explores the effectiveness question using Indiana vocational rehabilitation program as a case. Inspired by the literature on network effectiveness, this project evaluates PBC effectiveness from two perspectives: service outcome and participating organizations. Putting all the findings together, this project proposes that PBC seems more promising than FFS in human services. However, PBC effectiveness could not be well-rounded and should not be exaggerated. PBC, as a formal mechanism, adjusts contractor behavior through redefining incentive structure in formal contract design. Unfortunately, this formal effort of using PBC in vocational rehabilitation services was disturbed by the highly uncertain nature of employment services. Thus, there are only incomplete improvement in employment performance, mostly in targeted performance areas, and risks of contractor opportunistic behaviors.

Indeed, the research and the practice of PBC tend to ignore the relational face of contracting. Relational contracting as a social control system, using informal and normative mechanisms (largely represented by interorganizational trust) to eliminate interest and goal incongruence between contracting parties, tends to encourage appropriate behaviors that could lead to desirable collaborative outcomes. In this line of reasoning, this paper proposes the managerial implications that public managers

might consider when using PBC, such as ongoing collaboration and negotiation in contract planning and implementation, long-term or open-ended contracts, and professionalism.

In sum, this project represents the first attempt to systematically examine PBC effectiveness in human services. It shows the difficulties and dynamics of introducing performance management to human service contracting. For various political and pragmatic reasons, performance management is everywhere (Behn, 2003). However, largely due to human services' ambiguous performance and high provider discretion, PBC in human services are always at the risk of "rewarding A, while hoping for B." Therefore, the project reminds that the launch of PBC should be very deliberate and careful. The efforts of introducing PBC to human service provision are often undermined by imperfect performance measures and high provider discretion. The situation becomes worse-off when contractors use discretion to "gaming" the performance measures. Generally, in human services, not all aspects of performance can be clearly defined and measured. Along the full spectrum of the performance of a human service, there is some portion that is straightforward and easy to capture, such as successful placement and time-to-placement in this study. But there must be a certain portion, especially related to service quality and long-term effects, which is elusive to observe and define, such as positive quality-of-life change and long-term stability. In this way, the use of PBC with surrogate performance measures to adjust for the entire performance domain inevitably leads to a mismatch, ending up with incomplete performance improvement or even gaming (Dixit, 2002). Indeed, the

more discretion involved in human service delivery, the less portion of service performance can be clearly captured and measured (Lipsky, 1980). The broader ambiguous portion of service performance, the less effectiveness PBC could produce as a formal control mechanism, and the more room left for relational contracting to fit in. In conclusion, in order to take full advantage of PBC, public managers should pay attention to the relational side of contracting and devote administrative resources to building trust with contractors.

More broadly, the project underscores two key components of contracting management: (formal) control and trust. The issue of control is a lingering question in organizational management. Studies of management and organization behaviors have long examined effective ways to exercise control of collective actions (e.g., Barnard, 1938; Etzioni, 1964). In contracting management, control, as a power of directing, can be reflected in the provisions of the contract, monitoring, and levying of penalties. It includes monitoring of information flows, design of incentives, and allocation of risks. These formal mechanisms are absolutely necessary given the public accountability requirements. However, the effectiveness of such control is dependent upon the measurability of job-related behavior or outcome. When formal control systems are disturbed by various sociological and psychological factors, formal mechanisms become less effective and more costly. Indeed, due to information asymmetry and uncertainty, contracting out always features some ambiguity, even if in the areas other than human services. Government will never have complete access to, or influence over, contractors' operation and resources.

Thus, flexibility enjoyed by contractors is unavoidable. In this way, there must be areas that control mechanisms could not reach, but social control might emerge to play a role. Social control, based on social and normative influence, targets norms, values, and attitudes that may be relevant to desired collective outcomes (O'Reilly & Chatman, 1996). The center of social control is trust, which could create an environment where the mutually agreed contract goals become self-enforcing. This hybrid contracting management approach would help public managers address the smart-buy challenge and promote high-quality results.

Certainly, the arguments here are derived from the case study of Indiana vocational rehabilitation program. The external validity of a case study, as Yin (2009) suggests, lies in “analytical generalization” through replication rather than “statistical generalization” through inference from a sample to a population. This replication logic in theory testing and development demands that the robustness of a theory be confirmed only by replicating the findings in different contexts. In this sense, the research here represents one of the studies that systematically examines the effectiveness of PBC in human service provision. The findings here might be used only for conditional, contingent generalizations (George & Bennett, 2005) to other cases which are similar to the one under study. This project has no intention to generalize in order to infer the causal mechanisms under various contexts, although the findings here to some extent coincide with several recent studies in different human service areas (e.g., Heinrich & Choi, 2007; McGrew et al., 2005).

## Bibliography

- Abadie, A. (2005). Semiparametric Difference-In-Differences Estimators. *Review of Economic Studies*, 72(1), 1-19.
- Abadie, A., & Imbens, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1), 235-267.
- Ai, C., & Norton, E. C. (2003). Interaction Terms in Logit and Probit Models. *Economics letters*, 80(1), 123-129.
- Aldrich, H., & Herker, D. (1977). Boundary Spanning Roles and Organization Structure. *Academy of Management Review*, 2(2), 217-230.
- Amirkhanyan, A. A. (2010). Monitoring across Sectors: Examining the Effect of Nonprofit and For-Profit Contractor Ownership on Performance Monitoring in State and Local Contracts. *Public Administration Review*, 70(5), 742-755.
- Arrow, K. J. (1964). Control in Large Organizations. *Management Science*, 10(3), 397-408.
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Baker, G. P. (1992). Incentive Contracts and Performance Measurement. *Journal of Political Economy*, 100(3), 598-614.
- Baker, G. (2002). Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources*, 37(4), 728-751.
- Bardach, E. (1977). *The Implementation Game : What Happens after a Bill Becomes a Law*. Cambridge, MA: MIT Press.
- Barnard, C. (1938). *The Functions of the Executive*. Cambridge, MA: Harvard University Press.

- Barnow, B. S. (2000). Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act. *Journal of Policy Analysis and Management*, 19(1), 118-141.
- Becker, S. O., & Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *Stata Journal*, 2(4), 358-377.
- Behn, R. D. (2002). Government Performance and the Conundrum of Public Trust. In J. D. Donahue & J. S. Nye, Jr. (Eds.), *Market-based governance: Supply side, demand side, upside, and downside* (pp. 323-348). Washington, DC: Brookings Institution Press.
- Behn, R. D. (2003). Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*, 63(5), 586-606.
- Behn, R. D., & Kant, P. A. (1999). Strategies for Avoiding the Pitfalls of Performance Contracting. *Public Productivity & Management Review*, 22(4), 470-489.
- Beinecke, R. H., & DeFillippi, R. (1999). The Value of the Relationship Model of Contracting in Social Services Reprocurements and Transitions: Lessons from Massachusetts. *Public Productivity & Management Review*, 22(4), 490-501.
- Ben-Ner, A., Ren, T., & Paulson, D. F. (2011). A Sectoral Comparison of Wage Levels and Wage Inequality in Human Services Industries. *Nonprofit and Voluntary Sector Quarterly*, 40(4), 608-633.
- Berman, P. (1978). The study of macro- and micro- implementation. *Public Policy*, 26(2), 157-184.

- Bernheim, B. D., & Whinston, M. D. (1998). Incomplete Contracts and Strategic Ambiguity. *American Economic Review*, 88(4), 902-932.
- Bertelli, A. M., & Smith, C. R. (2010). Relational Contracting and Network Management. *Journal of Public Administration Research and Theory*, 20(suppl 1), i21-i40.
- Bevan, G., & Hood, C. (2006). What's Measured is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration*, 84(3), 517-538.
- Block, S. R., Athens, K., & Brandenburg, G. (2002). Using Performance-Based Contracts and Incentive Payments with Managed Care: Increasing Supported Employment Opportunities for People with Developmental Disabilities. *Journal of Vocational Rehabilitation*, 17(3), 165-174.
- Bohte, J., & Meier, K. J. (2000). Goal Displacement: Assessing the Motivation for Organizational Cheating. *Public Administration Review*, 60(2), 173-182.
- Bolton, B. F., Bellini, J. L., & Brookings, J. B. (2000). Predicting Client Employment Outcomes from Personal History, Functional Limitations, and Rehabilitation Services. *Rehabilitation Counseling Bulletin*, 44(1), 10-21.
- Bond, G. R. (2004). Supported Employment: Evidence for An Evidence-based Practice. *Psychiatric Rehabilitation Journal*, 27(4), 345-359.
- Boris, E. T., de Leon, E., Roeger, K. L., & Nikolova, M. (2010). *Human Service Nonprofits and Government Collaboration*. Washington, DC: Urban Institute.
- Brodkin, E. Z. (1997). Inside the Welfare Contract: Discretion and Accountability in State Welfare Administration. *Social Service Review*, 71:1-33.

- Brodkin, E. Z. (2011). Policy Work: Street-Level Organizations Under New Managerialism. *Journal of Public Administration Research and Theory*, 21(suppl 2), i253-i277.
- Brooke, V., Green, H., O'Brien, D., White, B., & Armstrong, A. (2000). Supported Employment: It's Working in Alabama. *Journal of Vocational Rehabilitation*, 14(3), 163-171.
- Brown, T. L., & Potoski, M. (2004). Managing the Public Service Market. *Public Administration Review*, 64(6), 656-668.
- Brudney, J. L., Fernandez, S., Ryu, J. E., & Wright, D. S. (2005). Exploring and Explaining Contracting Out: Patterns among the American States. *Journal of Public Administration Research and Theory*, 15(3), 393-419.
- Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1), 31-72.
- Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston: Houghton Mifflin.
- Chapin, J., & Fetter, B. (2002). Performance-based Contracting in Wisconsin Public Health: Transforming State-Local Relations. *Milbank Quarterly*, 80(1), 97-124.
- Cochran, W. G., & Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhyā: The Indian Journal of Statistics*, 35(4), 417-446.
- Commons, M., McGuire, T. G., & Riordan, M. H. (1997). Performance contracting for substance abuse treatment. *Health Services Research*, 32(5), 631-650.

- Cooper, P. J. (2003). *Governing by Contract: Challenges and Opportunities for Public Managers*. Washington, DC: CQ Press.
- Courty, P., & Marschke, G. (2004). An Empirical Investigation of Gaming Responses to Explicit Performance Incentives. *Journal of Labor Economics*, 22(1), 23-56.
- Cragg, M. (1997). Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act. *Journal of Law, Economics, and Organization*, 13(1), 147-168.
- D'Agostino, R. B., Jr. (1998). Propensity Score Methods for Bias Reduction in the Comparison of A Treatment to A Non-randomized Control Group. *Statistics in Medicine*, 17(19), 2265-2281.
- Daly, D., Tucker-Tatlow, J., & Gibson, C. (2004). *Innovations in Performance-Based Contracting*. San Diego, CA: Southern Area Consortium of Human Services.
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a Stewardship Theory of Management. *Academy of Management Review*, 22(1), 20-47.
- De Cooman, R., De Gieter, S., Pepermans, R., & Jegers, M. (2011). A Cross-sector Comparison of Motivation-related Concepts in For-profit and Not-for-profit Service Organizations. *Nonprofit and Voluntary Sector Quarterly*, 40(2), 296-317.
- Dehejia, R. H., & Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.

- Dehejia, R. H., & Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics*, 84(1), 151-161.
- DeHoog, R. H. (1984). *Contracting Out for Human Services: Economic, Political, and Organizational Perspectives*. Albany, NY: SUNY Press.
- DeHoog, R. H. (1990). Competition, Negotiation, or Cooperation: Three Models for Service Contracting. *Administration & Society*, 22(3), 317-340.
- Denzin, N. K. (1978). *The Research Act* (2nd ed.). New York: McGraw-Hill.
- Derthick, M. (1972). *New Towns In-town: Why a Federal Program Failed*. Washington, DC: The Urban Institute.
- DeVaro, J., & Brookshire, D. (2007). Promotions and Incentives in Nonprofit and For-profit Organizations. *Industrial and Labor Relations Review*, 311-339.
- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147-160.
- Dyer, J., H. Singh. (1998). The Relational View: Cooperative Strategy and Sources of interorganizational Competitive Advantage. *Academy of Management Review*, 23, 660-679.
- Dias, J. J., & Maynard-Moody, S. (2007). For-profit Welfare: Contracts, Conflicts, and the Performance Paradox. *Journal of Public Administration Research and Theory*, 17(2), 189-211.

- Dicke, L. A. (2002). Ensuring Accountability in Human Services Contracting Can Stewardship Theory Fill the Bill? *American Review of Public Administration*, 32(4), 455-470.
- Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources*, 37(4): 696-727.
- Donahue, J. D., & Nye, J. S. (Eds.). (2002). *Market-based Governance: Supply Side, Demand Side, Upside, and Downside*. Washington, DC: Brookings Institution Press.
- Dooley, D., Fielding, J., & Levi, L. (1996). Health and Unemployment. *Annual Review of Public Health*, 17(1), 449-465.
- Dutta, A., Gurvey, R., Chan, F., Chou, C.-C., & Ditchman, N. (2008). Vocational Rehabilitation Services and Employment Outcomes for People with Disabilities: A United States Study. *Journal of Occupational Rehabilitation*, 18(4), 326-334.
- Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *Academy of Management Review*, 14(1), 57-74.
- Elmore, R. F. (1979). Backward Mapping: Implementation Research and Policy Decisions. *Political Science Quarterly*, 94(4), 601-616.
- Ernita Joaquin, M., & Greitens, T. J. (2012). Contract Management Capacity Breakdown? An Analysis of U.S. Local Governments. *Public Administration Review*, 72(6), 807-816.
- Etzioni, A. (1964). *Modern Organizations*. Englewood Cliffs, NJ: Prentice-Hall.

- Faith, J., Panzarella, C., Spencer, R., Williams, C., Brewer, J., & Covone, M. (2010). Use of Performance-Based Contracting to Improve Effective Use of Resources for Publicly Funded Residential Services. *The Journal of Behavioral Health Services & Research*, 37(3), 400-408.
- Faems, D., Janssens, M., Madhok, A., & Van Looy, B. (2008). Toward an Integrative Perspective on Alliance Governance: Connecting Contract Design, Trust Dynamics, and Contract Application. *Academy of Management Journal*, 51(6), 1053-1078.
- Fawber, H. L., & Wachter, J. F. (1987). Job Placement as a Treatment Component of the Vocational Rehabilitation Process. *Journal of Head Trauma Rehabilitation*, 2(1), 27-33.
- Firestone, W. A. (1987). Meaning in Method: The Rhetoric of Quantitative and Qualitative Research. *Educational Researcher*, 16(7), 16-21.
- Frederickson, D. G., & Frederickson, H. G. (2006). *Measuring the Performance of the Hollow State*. Washington, D.C.: Georgetown University Press.
- Frumkin, P. (2001). *Managing outcomes: Milestone contracting in Oklahoma*. Washington, DC: The IBM Center for The Business of Government.
- Gamble, D., & Moore, C. L. (2003). The Relation between VR Services and Employment Outcomes of Individuals with Traumatic Brain Injury. *Journal of Rehabilitation*, 69(3), 31-38.
- Gates, L. B., Klein, S. W., Akabas, S. H., Myers, R., Schwager, M., & Kaelin-Kee, J. (2004). Performance-based contracting: turning vocational policy into jobs. *Administration and policy in mental health*, 31(3), 219-240.

- Gates, L. B., Klein, S. W., Akabas, S. H., Myers, R., Schwager, M., & Kaelin-Kee, J. (2004). Performance-based Contracting: Turning Vocational Policy into Jobs. *Administration and Policy in Mental Health, 31*(3), 219-240.
- Gaynor, M. (1990). Incentive Contracting in Mental Health: State and Local Relations. *Administration and Policy in Mental Health, 18*(1), 33-42.
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Ghoshal, S., & Moran, P. (1996). Bad for Practice: A Critique of the Transaction Cost Theory. *Academy of Management Review, 21*(1), 13-47.
- Giffords, E. D. (2003). An Examination of Organizational and Professional Commitment among Public, Not-For-Profit, and Proprietary Social Service Employees. *Administration in Social Work, 27*(3), 5-23.
- Girth, A. M., & Johnston, J. M. (2011). *Local Government Contracting*. National League of Cities.
- Glazerman, S., Levy, D. M., & Myers, D. (2003). Nonexperimental Versus Experimental Estimates of Earnings Impacts. *The Annals of the American Academy of Political and Social Science, 589*(1), 63-93.
- Glover, R. W., & Berger, B. L. (1989). Performance Contracting: The Colorado Model. *The Journal of Mental Health Administration, 16*(1), 21-28.
- Gramlich, E. M., & Koshel, P. P. (1975). *Educational Performance Contracting*. Washington, DC: Brookings Institution.
- Greene, W. (2010). Testing Hypotheses about Interaction Terms in Nonlinear Models. *Economics Letters, 107*(2), 291-296.

- Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal Multivariate Matching Before Randomization. *Biostatistics*, 5(2), 263-275.
- Gulati, R. (1995). Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances. *Academy of Management Journal*, 38(1), 85-112.
- Guo, S., & Fraser, M. W. (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications.
- Hart, O. (1989). An Economist's Perspective on the Theory of the Firm. *Columbia Law Review*, 1757-1774.
- Hart, O. D. (1988). Incomplete Contracts and the Theory of the Firm. *Journal of Law, Economics, & Organization*, 4(1), 119-139.
- Hasenfeld, Y. (1983). *Human Service Organizations*. Englewood Cliffs, NJ: Prentice-Hall.
- Hatry, H. P. (2006). *Performance Measurement: Getting Results*. Washington, DC: The Urban Insiteute.
- Haviland, A., Nagin, D. S., & Rosenbaum, P. R. (2007). Combining Propensity Score Matching and Group-Based Trajectory Analysis in An Observational Study. *Psychological Methods*, 12(3), 247-267.
- Heckman, J., Heinrich, C., & Smith, J. (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review*, 87(2), 389-395.
- Heckman, J., Heinrich, C., & Smith, J. (2003). Performance of Performance Standards. *Journal of Human Resources*, 37 (4), 778-811.

- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4), 605-654.
- Hefetz, A., & Warner, M. (2004). Privatization and Its Reverse: Explaining the Dynamics of the Government Contracting Process. *Journal of Public Administration Research and Theory*, 14(2), 171-190.
- Heinrich, C. J. (1999). Do Government Bureaucrats Make Effective Use of Performance Management Information? *Journal of Public Administration Research and Theory*, 9(3), 363-394.
- Heinrich, C. J. (2000). Organizational Form and Performance: An Empirical Investigation of Nonprofit and For-Profit Job-Training Service Providers. *Journal of Policy Analysis and Management*, 19(2), 233-261.
- Heinrich, C. J. (2002). Outcomes-based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness. *Public Administration Review*, 62(6), 712-725.
- Heinrich, C. J., & Choi, Y. (2007). Performance-Based Contracting in Social Welfare Programs. *American Review of Public Administration*, 37(4), 409-435.
- Heinrich, C. J., & Fournier, E. (2004). Dimensions of Publicness and Performance in Substance Abuse Treatment Organizations. *Journal of Policy Analysis and Management*, 23(1), 49-70.
- Heinrich, C. J., & Marschke, G. (2010). Incentives and their Dynamics in Public Sector Performance Management Systems. *Journal of Policy Analysis and Management*, 29(1), 183-208.

- Hill, C. J. (2006). Casework Job Design and Client Outcomes in Welfare-To-Work Offices. *Journal of Public Administration Research and Theory*, 16(2), 263-288.
- Hjern, B., & Porter, D. O. (1981). Implementation Structures: A New Unit of Administrative Analysis. *Organization Studies*, 2(3), 211-227.
- Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 15, 199–236.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945-960.
- Hood, C. (2006). Gaming in Targetworld: The Targets Approach to Managing British Public Services. *Public Administration Review*, 66(4), 515-521.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4), 305-360.
- Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), 602-611.
- Joaquin, M. E., & Greitens, T. J. (2012). Contract Management Capacity Breakdown? An Analysis of US Local Governments. *Public Administration Review*, 72(6), 807-816.
- Johnston, J. M., & Girth, A. M. (2012). Government Contracts and "Managing the Market": Exploring the Costs of Strategic Management Responses to Weak Vendor Competition. *Administration and Society*, 44(1), 3-29.

- Johnston, J. M., & Romzek, B. S. (1999). Contracting and Accountability in State Medicaid Reform: Rhetoric, Theories, and Reality. *Public Administration Review*, 59(5), 383-399.
- Joyce, P. G. (1993). Using Performance Measures for Federal Budgeting: Proposals and Prospects. *Public Budgeting & Finance*, 13(4), 3-17.
- Karaca-Mandic, P., Norton, E. C., & Dowd, B. (2012). Interaction Terms in Nonlinear Models. *Health Services Research*, 47(1), 255-274.
- Kaye, H. S. (1998). *Vocational Rehabilitation in the United States*. Washington, DC: National Institute on Disability and Rehabilitation Research (NIDRR).
- Kearney, K. A., McEwen, E., Bloom-Ellis, B., & Jordan, N. (2010). Performance-based Contracting in Residential Care and Treatment: Driving Policy and Practice Change through Public-Private Partnership in Illinois. *Child Welfare*, 89(2), 39-55.
- Keiser, L. R. (2010). Understanding Street-level Bureaucrats' Decision Making: Determining Eligibility in the Social Security Disability Program. *Public Administration Review*, 70(2), 247-257.
- Kelman, S. (2002). Strategic Contracting Management. In J. D. Donahue & J. S. Nye, Jr. (Eds.), *Market-based governance: Supply side, demand side, upside, and downside* (pp. 88-102). Washington, DC: Brookings Institution Press.
- Kerr, S. (1975). On the Folly of Rewarding A, While Hoping for B. *Academy of Management Journal*, 18(4), 769-783.
- Kettl, D. F. (1988). *Government by Proxy:(Mis?) Managing Federal Programs*. Washington, DC: CQ Press.

- Kettl, D. F. (1993). *Sharing Power: Public Governance and Private Markets*. Washington, DC: Brookings Institution Press.
- Kettl, D. F. (2002). *The Transformation of Governance: Public Administration for Twenty-First Century America*. Baltimore, MD: Johns Hopkins University Press.
- Kettl, D. F. (2002). Managing Indirect Government. In L. M. Salamon (Ed.), *The Tools of Government: A Guide to the New Governance* (pp. 490-510). New York: Oxford University Press.
- Kettl, D. F. (2005). *The Global Public Management Revolution*. Washington, DC: Brookings Institution Press.
- Kettner, P. M., & Martin, L. L. (1993). Performance, Accountability, and Purchase of Service Contracting. *Administration in Social Work, 17*(1), 61-79.
- Kim, Y. W., & Brown, T. L. (2012). The Importance of Contract Design. *Public Administration Review, 72*(5), 687-696.
- Kingdon, J. W. (1999). *America the Unusual*. Belmont, CA: Thomson/Wadsworth.
- Klein Woolthuis, R., Hillebrand, B., & Nooteboom, B. (2005). Trust, Contract and Relationship Development. *Organization Studies, 26*(6), 813-840.
- Koning, P., & Heinrich, C. J. (2013). Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts. *Journal of Policy Analysis and Management, 32*(3), 461-483.
- Krauskopf, J. (2008). *Performance Measurement in Human Services Contracts*. New York Nonprofit Press, 7(2).

- Kravchuk, R. S., & Schack, R. W. (1996). Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993. *Public Administration Review*, 56(4): 348-358.
- Lambright, K. T. (2009). Agency Theory and Beyond: Contracted Providers' Motivations to Properly Use Service Monitoring Tools. *Journal of Public Administration Research and Theory*, 19(2), 207-227.
- Lamothe, S., & Lamothe, M. (2012). Understanding the Differences between Vendor Types in Local Governance. *American Review of Public Administration*, 43(60), 709-728.
- Lamothe, M., & Lamothe, S. (2012). What Determines the Formal Versus Relational Nature of Local Government Contracting?. *Urban Affairs Review*, 48(3), 322-353.
- Leete, L. (2000). Wage Equity and Employee Motivation in Nonprofit and For-profit Organizations. *Journal of Economic Behavior & Organization*, 43(4), 423-446.
- Levine, D. M., American Educational Research, & American Association of School. (1972). *Performance Contracting in Education--An Appraisal: Toward A Balanced Perspective*. Englewood Cliffs, NJ: Educational Technology Publications.
- Lewis, J. D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, 63(4), 967-985.
- Linn, M. W., Sandifer, R., & Stein, S. (1985). Effects of Unemployment on Mental and Physical Health. *American Journal of Public Health*, 75(5), 502-506.

- Lipsky, M. (1980). *Street-level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Lyons, B., & Mehta, J. (1997). Contracts, Opportunism and Trust: Self-interest and Social Orientation. *Cambridge Journal of Economics*, 21(2), 239-257.
- Lu, M. (1999). Separating the True Effect from Gaming in Incentive-Based Contracts in Health Care. *Journal of Economics and Management Strategy*, 8(3), 383–431.
- Luhmann, N. 1979: *Trust and Power*. Chichester: Wiley.
- Lunceford, J. K., & Davidian, M. (2004). Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study. *Statistics in Medicine*, 23(19), 2937-2960.
- Macaulay, S. (1963). Non-contractual Relations in Business: A Preliminary Study. *American Sociological Review*, 28(1), 55-67.
- Macaulay, S. (1985). An Empirical View of Contract. *Wisconsin Law Review*, 5, 465-482.
- Macneil, I. R. (1977). Contracts: Adjustment of Long-Term Economic Relations under Classical, Neoclassical, and Relational Contract Law. *Northwestern University Law*, 72, 854-902.
- Macneil, I. R. (1980). *The New Social Contract: An Inquiry into Modern Contractual Relations*. New Haven, CT: Yale University Press.
- Mathison, S. (1988). Why Triangulate?. *Educational Researcher*, 17(2), 13-17.
- Martin, L. L. (1999). Performance Contracting: Extending Performance Measurement to Another Level. *Public Administration Times*, 22 (January): 1 & 2.

- Martin, L. L. (2005). Performance-based Contracting for Human Services: Does it Work?. *Administration in Social Work, 29*(1), 63-77.
- Martin, L. L., & Kettner, P. M. (1996). *Measuring the Performance of Human Service Programs* Thousand Oaks, CA: Sage.
- Marvel, M. K., & Marvel, H. P. (2007). Outsourcing Oversight: A Comparison of Monitoring for In-house and Contracted Services. *Public Administration Review, 67*(3), 521-530.
- Matland, R. E. (1995). Synthesizing the Implementation Literature: The Ambiguity-conflict Model of Policy Implementation. *Journal of Public Administration Research and Theory, 5*(2), 145-174.
- Mazmanian, D. A., & Sabatier, P. A. (1983). *Implementation and Public Policy*. Glenview, IL: Scott Foresman.
- McEvily, B., Perrone, V., & Zaheer, A. (2003). Trust as An Organizing Principle. *Organization Science, 14*(1), 91-103.
- McGrew, J. H., Johannesen, J. K., Griss, M. E., Born, D. L., & Katuin, C. (2005). Performance-based Funding of Supported Employment: A Multi-site Controlled Trial. *Journal of Vocational Rehabilitation, 23*(2), 81-99.
- McGrew, J., Johannesen, J., Griss, M., Born, D., & Katuin, C. (2007). Performance-based Funding of Supported Employment for Persons with Severe Mental Illness: Vocational Rehabilitation and Employment Staff Perspectives. *The Journal of Behavioral Health Services and Research, 34*(1), 1-16.

- McLellan, A. T., Kemp, J., Brooks, A., & Carise, D. (2008). Improving Public Addiction Treatment through Performance Contracting: The Delaware Experiment. *Health Policy*, 87(3), 296-308.
- Mecklenburger, J. (1972). *Performance Contracting*. Worthington, OH: C.A. Jones.
- Meyers, M. K., Glaser, B., & Donald, K. M. (1998). On the Front Lines of Welfare Delivery: Are Workers Implementing Policy Reforms? *Journal of Policy Analysis and Management*, 17(1), 1-22.
- Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? *Review of Economics and Statistics*, 86(1), 156-179.
- Milgrom, P., & Roberts, J. (1992). *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Miller, S., & Wilson, N. (1981). The Case for Performance Contracting. *Administration and Policy in Mental Health and Mental Health Services Research*, 8(3), 185-193.
- Milward, H. B., & Provan, K. G. (2000). Governing the Hollow State. *Journal of Public Administration Research and Theory*, 10(2), 359-380.
- Moynihan, D. P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Novak, J., Mank, D., Revell, G., & Zemaitis, N. (1999). Initiatives Influencing the Emergence of Results-based Funding of Supported Employment Services. In g. Revell, K. J. Inge, D. Mank, & P. Wehman (Eds.), *The Impact of Supported*

- Employment for People with Significant Disabilities (pp. 25-42). Richmond, VA: Virginia Commonwealth University, Rehabilitation Research & Training Center on Workplace Supports.
- O'Brien, D., & Revell, G. (2005). The Milestone Payment System: Results-based Funding in Vocational Rehabilitation - 2005. *Journal of Vocational Rehabilitation, 23*(2), 101-114.
- O'Brien, D., & Revell, G. (2006). Current Trends in Funding Employment Outcomes. In Wehman, P., Inge, K. J., Revell, G., & Brooke, V. A. (Eds.) *Real Work for Real Pay: Inclusive Employment for People with Disabilities*. Baltimore, MD: Paul Brookes Publishing.
- Okun, A. M. (1975). *Equality and Efficiency: The Big Tradeoff*. Washington, DC: Brookings Institution Press.
- O'Reilly, C. A., & Chatman, J. A. (1996). Culture as Social Control: Corporations, Cults, and Commitment. *Research in Organizational Behavior, 18*(18), 157-200.
- Osborne, D., & Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Reading, MA: Addison-Wesley.
- Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action: *American Political Science Review, 92*(1), 1-22.
- O'Toole, L. J. (2000). Research on Policy Implementation: Assessment and Prospects. *Journal of Public Administration Research and Theory, 10*(2), 263-288.

- Ouchi, W. G. (1980). Markets, Bureaucracies, and Clans. *Administrative Science Quarterly*, 25(1), 129-141.
- Ouchi, W. G., & Maguire, M. A. (1975). Organizational Control: Two Functions. *Administrative Science Quarterly*, 20(4), 559-569.
- Paul, K. I., & Moser, K. (2009). Unemployment Impairs Mental Health: Meta-analyses. *Journal of Vocational Behavior*, 74(3), 264-282.
- Poppo, L., & Zenger, T. (2002). Do Formal Contracts and Relational Governance Function as Substitutes or Complements?. *Strategic Management Journal*, 23(8), 707-725.
- Pressman, J. L., & Wildavsky, A. (1984). Implementation. Univ of California Press.
- Prottas, J. M. (1978). The Power of the Street-Level Bureaucrat in Public Service Bureaucracies. *Urban Affairs Review*, 13(3), 285-312.
- Provan, K. G., & Milward, H. B. (2001). Do Networks Really Work? A Framework for Evaluating Public-Sector Organizational Networks. *Public Administration Review*, 61(4), 414-423.
- Puhani, P. A. (2012). The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-In-Differences" Models. *Economics Letters*, 115(1), 85-87.
- Radin, B. (2006). *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.
- Revell, W. G., West, M., & Cheng, Y. (1998). Funding Supported Employment: Are There Better Ways?. *Journal of Disability Policy Studies*, 9(1), 59-79.

- Riccucci, N. (2005). *How Management Matters: Street-level Bureaucrats and Welfare Reform*. Washington, DC: Georgetown University Press.
- Ring, P. S., & Van de Ven, A. H. (1994). Developmental Processes of Cooperative Interorganizational Relationships. *Academy of Management Review*, 19(1), 90-118.
- Romzek, B. S. (2000). Dynamics of Public Sector Accountability in An Era of Reform. *International Review of Administrative Sciences*, 66(1), 21-44.
- Romzek, B. S., & Johnston, J. M. (2002). Effective Contract Implementation and Management: A Preliminary Model. *Journal of Public Administration Research and Theory*, 12(3), 423-453.
- Romzek, B. S., & Johnston, J. M. (2005). State Social Services Contracting: Exploring the Determinants of Effective Contract Accountability. *Public Administration Review*, 65(4), 436-449.
- Romzek, B. S., LeRoux, K., & Blackmar, J. M. (2012). A Preliminary Theory of Informal Accountability among Network Organizational Actors. *Public Administration Review*, 72 (3), 442-453.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *American Statistician*, 39(1), 33-38.

- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All: A Cross-discipline View of Trust. *Academy of Management Review*, 23(3), 393-404.
- Rubin, D. B. (1973). Matching to Remove Bias in Observational Studies. *Biometrics*, 29(1), 159-183.
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581-592.
- Rubin, D. B. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association*, 74(366), 318-328.
- Rubin, D. B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Annals of Internal Medicine*, 127(8), 757-763.
- Rubin, D. B., & Thomas, N. (1996). Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*, 52(1), 249-264.
- Rubin, S. E., Roessler, R., & Dunkerby, M. (1983). *Foundations of the Vocational Rehabilitation Process*. Boston: University Park Press.
- Sabatier, P. A. (1986). Top-Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis. *Journal of Public Policy*, 6(01), 21-48.
- Salamon, L. M. (1987). Of Market Failure, Voluntary Failure, and Third-Party Government: Toward a Theory of Government-Nonprofit Relations in the Modern Welfare State. *Nonprofit and Voluntary Sector Quarterly*, 16(1-2), 29-49.

- Salamon, L. M. (1995). *Partners in Public Service: Government-Nonprofit Relations in the Modern Welfare State*. Baltimore, MD: Johns Hopkins University Press.
- Salamon, L. M. (1989). *Beyond Privatization: The Tools of Government Action*. Washington, DC: Urban Institute Press.
- Sandfort, J. R. (2000). Moving beyond Discretion and Outcomes: Examining Public Management from the Front Lines of The Welfare System. *Journal of Public Administration Research and Theory*, 10(4), 729-756.
- Savas, E. S. (1987). *Privatization: The Key to Better Government*. Chatham, N.J.: Chatham House.
- Schlesinger, M., Dorwart, R. A., & Pulice, R. T. (1986). Competitive Bidding and States' Purchase of Services: The Case of Mental Health Care in Massachusetts. *Journal of Policy Analysis and Management*, 5(2), 245-263.
- Schlesinger, M., Mitchell, S., & Gray, B. H. (2004). Public Expectations of Nonprofit and For-profit Ownership in American Medicine: Clarifications and Implications. *Health Affairs*, 23(6), 181-191.
- Sclar, E. D. (2001). *You Don't Always Get What You Pay for: The Economics of Privatization*. Ithaca, NY: Cornell University Press.
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. *Journal of the American Statistical Association*, 103(484), 1334-1343.

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Independence, KY: Wadsworth Cengage learning.
- Shapiro, S. P. (2005). Agency Theory. *Annual Review of Sociology*, 31, 263-284.
- Smith, D. C., & Grinker, W.J. (2004). The Promise and Pitfalls Of Performance-Based Contracting. Presentation at the 25th Annual Research Conference of the Association for Public Policy Analysis and Management (APPAM), Washington, DC. November 5-8, 2003.
- Smith, S. R., & Smyth, J. (1996). Contracting for Services in a Decentralized System. *Journal of Public Administration Research and Theory*, 6(2), 277-296.
- Steckler, A., McLeroy, K. R., Goodman, R. M., & Bird, S. T. (1992). Toward Integrating Qualitative and Quantitative Methods: An Introduction. *Health Education Quarterly*, 19(1), 1-8.
- Stewart, M. T., Horgan, C. M., Garnick, D. W., Ritter, G., & McLellan, A. T. (2013). Performance Contracting and Quality Improvement in Outpatient Treatment: Effects on Waiting Time and Length of Stay. *Journal of Substance Abuse Treatment*, 44(1), 27-33.
- Stillman, R. J. (1991). Preface to Public Administration: A Search for Themes and Direction. New York: St. Martin's Press.
- Thompson, J. D. (1967). *Organisation in Action*. New York: MacGrow-Hill.
- Reuer, J. J., & Ariño, A. (2007). Strategic Alliance Contracts: Dimensions and Determinants of Contractual Complexity. *Strategic Management Journal*, 28(3), 313-330.

- U.S. Child Care Bureau. (2008). Child Care and Development Fund: Report of state and territory plans FY 2008-2009.
- U.S. Child Care Bureau. (2009). Examples of Performance-based Contracts in Child Welfare Services.
- U.S. Department of Education. (2010). RSA Annual Report for fiscal year 2010.
- U.S. Government Accountability Office (GAO). (2001). Contract Management: Trends and Challenges in Acquiring Services. GAO-01-753T.
- U.S. Government Accountability Office (GAO). (2002). Contract Management: Guidance Needed for Using Performance-Based Service Contracting. GAO-02-1049.
- U.S. Office of Federal Procurement Policy (OFPP). (2007). Fiscal Year 2008 Performance-Based Acquisition Performance Goal.
- U.S. Office of Federal Procurement Policy (OFPP). (2007). Using Performance-Based Acquisition to Meet Program Needs - Performance Goals, Guidance, and Training.
- Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, 42(1), 35-67.
- Van Slyke, D. M. (2003). The Mythology of Privatization in Contracting for Social Services. *Public Administration Review*, 63(3), 296-315.
- Van Slyke, D. M. (2007). Agents or Stewards: Using Theory to Understand the Government-Nonprofit Social Service Contracting Relationship. *Journal of Public Administration Research & Theory*, 17(2), 157-187.

- Van Thiel, S., & Leeuw, F. L. (2002). The Performance Paradox in the Public Sector. *Public Performance & Management Review*, 25(3), 267-281.
- Vandaele, D., Rangarajan, D., Gemmel, P., & Lievens, A. (2007). How to Govern Business Services Exchanges: Contractual and Relational Issues. *International Journal of Management Reviews*, 9(3), 237-258.
- Waernbaum, I. (2010). Propensity Score Model Specification for Estimation of Average Treatment Effects. *Journal of Statistical Planning and Inference*, 140(7), 1948-1956.
- Warner, M. E., & Hefetz, A. (2008). Managing Markets for Public Service: The Role of Mixed Public–Private Delivery of City Services. *Public Administration Review*, 68(1), 155-166.
- Warner, M. E., & Hefetz, A. (2009). Cooperative Competition: Alternative Service Delivery, 2002-2007. In *The Municipal Year Book 2009*, ed. ICMA, 11–20. Washington, DC: International City County Management Association.
- Wedel, K. R., & Conston, S. W. (1988). Performance Contracting for Human Services: Issues and Suggestions. *Administration in Social Work*, 12(1), 73-87.
- Williams, D. W. (2003). Measuring Government in the Early Twentieth Century. *Public Administration Review*, 63(6), 643-659.
- Williamson, O. E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.
- Wilson, J. Q. (2000). *Bureaucracy: What Government Agencies Do and Why They Do it*. New York: Basic Books.

- Witesman E. M., & Fernandez, S. (2013). Government Contracts With Private Organizations: Are There Differences Between Nonprofits and For-profits? *Nonprofit and Voluntary Sector Quarterly*, 42(4), 689-715.
- Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th ed.). Los Angeles, CA: Sage.
- Zand, D.E. (1972). Trust and Managerial Problem Solving. *Administrative Science Quarterly*, 17 (2), 229-239.
- Zhao, Z. (2008). Sensitivity of Propensity Score Methods to the Specifications. *Economics Letters*, 98(3), 309-319.
- Zollo, M., Reuer, J. J., & Singh, H. (2002). Interorganizational Routines and Performance in Strategic Alliances. *Organization Science*, 13(6), 701-713.
- Zucker, L. G. (1986). Production of Trust: Institutional Sources of Economic Structure, 1840–1920. *Research in Organizational Behavior*, 8, 53-111.