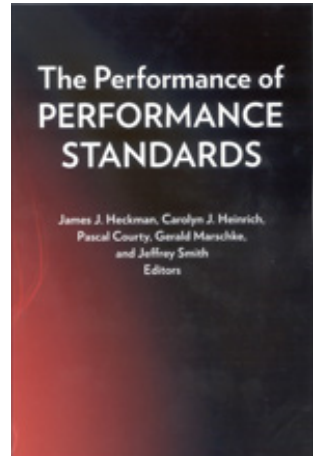

Upjohn Institute Press

Performance Standards and the Potential to Improve Government Performance

James J. Heckman
University of Chicago

Carolyn J. Heinrich
University of Wisconsin

Jeffrey Smith
University of Michigan



Chapter 1 (pp. 1-14) in:

The Performance of Performance Standards

James J. Heckman, Carolyn J. Heinrich, Pascal Courty, Gerald Marschke,
and Jeffrey Smith, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011

1

Performance Standards and the Potential to Improve Government Performance

James J. Heckman
Carolyn J. Heinrich
Jeffrey Smith

A number of recently implemented reforms to public sector incentive systems have sought to reorient them toward a focus on measuring *results* and inducing public agencies to become more efficient, responsive, and accountable to the public. They share at least two features. The first is a system of performance measures and standards designed to create clear expectations for government performance and to assess results. A second feature is a means for rewarding individuals, teams, or entire organizations for achievement relative to the established performance goals, primarily through budgetary allocations. By clearly defining goals and developing explicit rewards for their attainment, these systems have aimed to replicate, in a nonmarket setting, the incentive structures, competition, and resulting high performance and efficiency of private markets (Light 2005).

Among its first steps in advancing these reforms, the Obama administration has required federal agencies to identify a limited number of high-priority performance goals for which performance trends will be tracked, and through its new Open Government initiative, it will make these data publicly available and promote the use of new methods in the analysis of them. The government is also now compelling the private sector to provide more information on its performance for transparency and accountability (beyond longer-standing areas of public scrutiny such as health care and the environment), and has devised incentives for cooperation and penalties for withholding information (Cukier 2010). For example, it is now possible for the public to get sta-

tistics on job-related deaths that name employers and to see restaurants' health inspection scores online.

As the use of performance measurement and incentive systems has expanded in the public sector, so has the number of studies calling attention to their problems and unintended effects (Bevan and Hood 2006; Brooks 2000; Courty and Marschke 2004; Heckman and Smith 2004; Heinrich 2004, 2007; Heinrich and Marschke 2010; Koning and Heinrich 2010; Propper and Wilson 2003; Radin 2000; Smith 1995; General Accounting Office [GAO] 2002). Performance standards systems and bonuses are (or have been) used in Food Stamps (now the Supplemental Nutrition Assistance Program) and welfare-to-work programs, employment and training programs, public school accountability systems under No Child Left Behind, child welfare agencies and child support enforcement programs, Medicaid and SCHIP programs, and other social programs, although not without some degree of controversy and ongoing challenges in their design and implementation. The development of performance incentive systems in public bureaucracies also continues to advance in Europe, led by Great Britain's early exploration, and with some governments (such as Australia and the Netherlands) now implementing incentive systems with fully (100 percent) performance-contingent pay/contracting arrangements (Finn 2008; Struyven and Steurs 2005).

While the broad introduction of incentive systems in many government agencies is new, U.S. employment and training programs have used both performance standards and monetary bonuses for over two decades. Klerman (2005, p. 347) describes the Job Training Partnership Act (JTPA)/Workforce Investment Act (WIA) performance measurement system as one of the "most mature implementations of performance-based management." It is also one of the most studied systems, in part because of the randomized experimental evaluation of the JTPA program that produced important information for assessing the performance of these performance standards systems in measuring program impacts (Bloom et al. 1993; Dickinson et al. 1988; Heckman, Heinrich, and Smith 2002; Heckman, LaLonde, and Smith 1999; Orr et al. 1995). Policymakers have looked to the results of these studies to guide changes to these systems in employment and training programs and to also inform the design and operation of performance standards systems in other government programs.

At the same time, one of our motivations for assembling the research presented in this book is that despite decades of study and practice, some of the important lessons that have been learned are not reflected in the current design and implementation of performance standards systems. Bevan and Hood (2006), for example, describe the development and use of performance targets in the English public health care system, along with the perverse incentives they generated, as “hitting the target and missing the point.”¹ And despite the long tenure of performance standards in U.S. employment and training programs, a 2002 GAO report (p. 14) confirms the persistence of gaming responses that influence who gets access to program services as well as service intensity, concluding that “the need to meet performance levels may be the driving factor in deciding who receives WIA-funded services at the local level.”

It may be that some of the more rigorous empirical evidence from past studies has not been effectively communicated or penetrated policymaking and public management circles deeply enough.² Or, as Heckman, Heinrich, and Smith comment in Chapter 3, it may be that policymakers who have mandated such systems (and administrators involved in their implementation) have not fully appreciated the challenges of designing a performance management system that generates incentives for improving program impacts. Research in this area has also continued to evolve, with advances in theoretical conceptions, modeling, and data, and of course, changes in the incentive systems over time have facilitated analysis of their implications for performance standards system design and functioning. The chapters included in this volume embody a number of these important advancements, and yet our primary aim is to make the lessons of our research clear to those who design and implement these incentive systems in the public sector.

The nine chapters that follow use U.S. employment and training programs as a “laboratory” for investigation. They draw extensively from the data and experiences of the earlier JTPA performance standards system, both because of its longer tenure and the availability of experimental data. Using a variety of data sources on these incentive systems, the authors of these chapters explore how performance standards and incentives affect the behavior of public managers and agency employees, their approaches to service delivery, and ultimately, the outcomes for participants. Both the JTPA and WIA programs have allowed state and local administrators and their governing boards substantial

discretion, within broad limits, to determine performance goals, standards, and bonus systems. This administrative flexibility is reflected in the range of incentive systems that have been implemented by different states and by the same states over time. It is this variation that serves as the grist for our empirical mill and is used to extract general lessons that can be applied on a wider scale to both existing and newly developing performance incentive systems.

KEY QUESTIONS AND CHARACTERISTICS OF PERFORMANCE STANDARDS SYSTEMS IN U.S. EMPLOYMENT AND TRAINING PROGRAMS

While investigating formal incentive structures and organizational behavior within U.S. employment and training programs, the authors in this volume address the following six fundamental questions:

- 1) How do performance standards and measures operate to include or exclude individuals with different characteristics in these programs?
- 2) How do performance standards and measures affect the types of services offered?
- 3) How do the processes for setting standards and weights for performance goals and for recognizing and rewarding performance affect system incentives and bureaucratic responses?
- 4) Are the performance standards, measures, and incentives effective in motivating bureaucratic behavior toward the achievement of program goals?
- 5) Do short-term outcome measures used in the performance standards systems predict long-term impacts of the programs on participants?
- 6) What general lessons can be learned from a study of these performance standards systems and the variation in the rules governing their administration over time?

In the next chapter, the JTPA and WIA performance standards systems are described in greater detail. Below, we briefly list four key features of the programs that are pertinent to understanding the contributions of this volume. We follow with an overview of the chapters and their salient findings.

- 1) Federal job training programs (formerly JTPA and currently WIA) are highly decentralized. Local level governing boards include local business officials along with other government, community, and labor leaders. States control most aspects of the performance standards system that guides training activities, including determining the rules that govern how outcome measures count and how performance awards are made.
- 2) Workforce development programs generate relatively easily measured outcomes such as employment, wages, completion of education programs, enrollment in the military, and continued schooling. More difficult to assess, however, is the value added produced by the programs, which, as we show, does not necessarily correlate strongly with shorter-term outcome measures.
- 3) Most performance standards have been formulated in terms of *levels* of achievement, rather than in terms of achievement *gains* resulting from program participation. This practice gives rise to the potential for “cream skimming,” which results when centers enroll persons who are likely to have high outcome levels rather than those whose outcomes would improve most through participation in the program.
- 4) Training centers (or states under WIA) that meet or exceed their performance standards typically receive budgetary rewards. Training centers have shown, however, that they can manipulate the performance standards system in ways that improve measured performance but may not contribute to value-added or individual gains. The ability of program managers and staff to manipulate the data used to monitor them poses a major challenge to the successful design of performance standards systems.

OVERVIEW OF THE CHAPTERS AND THEIR RELATIONSHIP TO PREVIOUS RESEARCH

The studies in this volume take advantage of data that are superior in scope and detail to some of the data used in the existing literature or available on a regular basis for assessing program performance. The research described in Chapters 4, 6, 8, and 9 benefitted from detailed longitudinal, microlevel data that were collected in the National JTPA Study (NJS) and through other administrative data sources. The NJS data also include information on JTPA-eligible nonparticipants in four experimental sites. These data have two important advantages over samples of eligible populations constructed from large public-use data sets: 1) they are collected from eligible nonparticipants residing in the same geographical area as participants, and 2) the data are gathered using the same survey instruments as those used for participants.

In addition, the authors collect and analyze complete information about state-level variation in the JTPA and WIA performance standards systems. Chapters 4 and 5 show that state incentive systems are highly complex and differ widely across states and over time within states. They are not easily characterized by small dimensional summary measures as used in previous studies (see, for example, Anderson, Burkhauser, and Raymond [1993] and Dickinson et al. [1988]). This wealth of data is very helpful in assessing the implications of changes and differences in performance standard regimes, and a variety of analytical strategies are applied to learn from these data.

Before addressing the key questions listed above, the contributing authors provide some basic information in Chapter 2 about the design and implementation of the JTPA and WIA programs to aid our readers' understanding of the research that follows. Chapter 2 describes the origins and organizational structure of these programs, eligibility rules and the types of services made available to participants, details of the performance measures, and other aspects of these performance standards systems' design. Chapter 2 also highlights some of the changes in these systems over time, particularly in the shift from JTPA to WIA.

In Chapter 3, James Heckman, Carolyn Heinrich, and Jeffrey Smith set up a formal model of a performance standards system that theoretically frames many of the fundamental questions and issues that are

addressed in this book. For example, their model demonstrates how the JTPA and WIA program goals to promote equity and efficiency interact with a performance standards system based on short-term outcome measures. As such, they show that whether or not cream skimming is at odds with equity or efficiency (or both) is an empirical question that depends on the relationship between the benefits of program services and the location of trainees in the skill distribution of applicants. Importantly, the broad conceptual framework that they develop provides a foundation for future investigation about when (and if) performance standards systems in active labor market programs will increase labor market impacts, and when (and if) they are more likely to have unintended consequences due to responses by program staff to the incentives they provide.

In Chapter 4, Courty and Marschke describe the structure of the JTPA performance standards system and highlight important details that go into developing a performance-contingent budgeting scheme. They describe the performance measures upon which awards were contingent, the nature of the awards, and how performance awards were computed in JTPA; they also show how crucial the handling of these details can be to the success of performance funding. Courty and Marschke find that interstate differences in the implementation of the performance incentives system grew over time, reflecting the substantial discretion federal authorities gave to state and local agencies in its management. Their findings suggest that the objectives of states' incentive systems increasingly diverged from the original federal ones, leading to some unintended consequences.

Courty, Heinrich, and Marschke draw from the information economics, contract theory, and public administration literatures to derive theoretical implications for the establishment of appropriate benchmark levels of performance (i.e., performance standards) in Chapter 5. They then assess alternative methods that are commonly used to construct performance standards and evaluate their application in the JTPA and WIA systems. They find evidence that performance measurement system designers have attempted to "level the playing field" over time to provide equivalent performance incentives across states and localities. However, they also identify some negative dynamic properties of the JTPA and WIA performance measurement systems. The dynamics of performance benchmarking, and the politically motivated ambition to

demonstrate continuous performance improvement, likely introduced inefficiencies and generated incentives to influence performance in ways other than increasing effort. These potentially harmful behaviors include selecting trainees according to observed characteristics associated with their labor market success, limiting the availability of more intensive training services and demonstrating lower performance early on to allow for performance improvements over time.

In Chapter 6, Heckman and Smith present a detailed, step-by-step analysis of the process by which individuals become participants in employment and training programs. Using data from the JTPA system, the goal of their analysis is to determine the sources of demographic disparities in participation rates and to isolate the roles of personal information, program eligibility rules, and administrative discretion in determining the probability of program participation. Disparities arising from voluntary decisions not to participate by informed persons raise different policy concerns from disparities arising from inequitable administrative practices or bureaucratic treatment. Models for the determination of eligibility, awareness of the program, and application and acceptance into the program are estimated. Bureaucratic discretion is assumed to play the major role in the final stage of the process. The evidence reported in this chapter indicates that while cream skimming is an important feature of the JTPA program, program eligibility rules, the personal choices of potential participants, and informational constraints are also important determinants of demographic disparities in JTPA participation.

Another important lesson from the incentive literature is that explicit incentives may elicit unintended and dysfunctional responses, also known as gaming responses. In Chapter 7, Courty and Marschke present a comprehensive overview of the literature on dysfunctional responses and develop a theoretical framework to classify the various dysfunctional responses that have been identified in practice. They distinguish three types of dysfunctional responses: 1) accounting manipulations that have no impact on the organization, 2) gaming responses that boost performance outcomes but have a negative impact on the organization, and 3) marginal misallocations that have positive impacts but are suboptimal because alternative allocations would have a higher impact. They then summarize the empirical evidence of dysfunctional responses in the JTPA system and assess the extent to which

such responses may impede the effective functioning of performance measurement systems.

A widely cited GAO report (1991) claims that the JTPA performance standard system produced inequities by encouraging frequent routing of minorities and women into the least productive training activities and by denying them training altogether. The GAO alleges that training centers' pursuit of incentive payments and recognition for high performance has a perverse effect on participant selection as well as on trainees' access to services. In Chapter 8, through an extensive analysis of a JTPA training center in Illinois, Heinrich shows that bureaucratic decisions involving the selection of participants and their assignments to JTPA program services are frequently interdependent. She combines case-study and econometric approaches to investigate the screening decisions made by program staff and the influence of performance standards on these decisions. Heinrich finds evidence that these frontline staff cream-skim in making their enrollment and treatment-assignment decisions, and she also documents how shifts in federal policy fostered a movement away from more costly but higher value-added services toward less costly services. In the face of declining JTPA budgets, it appears that bureaucrats sought to preserve their client load and maintain low costs by offering cheaper but less cost-effective services. The resulting effect, most likely unintended, is that bureaucrats cut service quality to keep program costs low and maintain service quantity.

In Chapter 9, Heckman, Heinrich, and Smith use data from the NJS to examine the relationship between the short-run performance measures used in the JTPA performance standards system, such as employment and wages at termination from the program (or 13 weeks after termination), and experimental estimates of the impact or value added of the program on employment and earnings. In most cases, the data reveal a very weak relationship between the short-run measures and the long-run impacts. In fact, this relationship is often perverse, so that higher short-run measures are associated with lower long-term impacts from training. The JTPA performance standards system was, and the current WIA system is, based on measures that are not highly correlated with the gains from the program. Thus, evidence presented in the other chapters that the performance standards system "works" in the sense that it motivates bureaucrats to achieve certain standards does not imply that the performance standards system leads training centers to choose

the combination of persons served and services offered that maximizes total gains obtained from the program.

In Chapter 10, Heckman, Heinrich, and Smith summarize three main lessons of the volume regarding the effectiveness of performance standards in public organizations and the need for further research. First, organizations respond to incentives, but sometimes the responses are perverse. In the first iteration of an incentive system's design, well-meaning designers of the performance standards system are unlikely to anticipate the shrewd responses of program administrators and frontline workers to system incentives, or the many possible ways they might influence measured performance without necessarily adding to (or possibly even detracting from) program value or impact. For example, individuals' access to program services and the nature and duration of the services participants receive are sometimes adversely affected by bureaucrats' responses to the performance incentives.

Second, the short-term outcome measures that continue to be used in the WIA performance standards system are only weakly related to the true long-run impacts of the program on earnings and employment. Measures of changes in earnings were tried under WIA but were discontinued, and the U.S. Department of Labor (USDOL) continues with measures of entry into unsubsidized employment, retention in unsubsidized employment six months after entry into the employment, and earnings levels six months after entry into the employment. Researchers and policymakers have yet to identify performance measures that will promote key, long-term program objectives while simultaneously generating more readily available performance information for ongoing program management.

Third, the importance of the cream skimming issue has been overstated in popular discussions. In the provision of employment and training services, the trade-off between efficiency and equity is modest at best. Personal choices and informational constraints play an important role in accounting for demographic differences in program participation among those eligible, while the effect of administrative discretion in accounting for demographic disparities among the eligible population is relatively limited. Chapter 10 includes a brief discussion of some policy implications of these findings.

In his book *The Dynamics of Performance Management*, Donald Moynihan (2008, pp. 4–5) argues that performance management re-

forms have become so central to recent public management agendas that it “is only a slight exaggeration to say that we are betting the future of governance on the use of performance information.” Clearly, the design and implementation of performance standards and incentive systems in the public sector will continue to be a dynamic pursuit, and it is our hope that the lessons distilled in this volume will have a role in shaping and speeding their evolution and the improvement of government performance.

Notes

1. In the effort to achieve hospital accident and emergency waiting time targets, hospitals cancelled operations and required patients to wait in ambulance queues outside the hospital until they were confident the patients could be seen within the targeted (four-hour) time. Bevan and Hood also find discrepancies between official reported levels of performance and those from independent surveys of patients.
2. Or, it may be the case that some of the fundamental lessons from studies to date have been ignored or deferred in pursuit of other objectives (political or otherwise); see, for example, Radin (2000).

References

- Anderson, Kathryn, Richard Burkhauser, and J. Raymond. 1993. “The Effect of Creaming on Placement Rates under the Job Training Partnership Act.” *Industrial and Labor Relations Review* 46(4): 613–624.
- Bevan, Gwyn, and Christopher Hood. 2006. “What’s Measured Is What Matters: Targets and Gaming in the English Public Health Care System.” *Public Administration* 84(3): 517–538.
- Bloom, Howard, Larry Orr, George Cave, Steve Bell, and Fred Doolittle. 1993. *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months*. Bethesda, MD: Abt Associates.
- Brooks, Arthur C. 2000. “The Use and Misuse of Adjusted Performance Measures.” *Journal of Policy Analysis Management* 19(2): 323–329.
- Courty, Pascal, and Gerald Marschke. 2004. “An Empirical Investigation of Gaming Responses to Performance Incentives.” *Journal of Labor Economics* 22(1): 23–56.
- Cukier, Kenneth. 2010. “Data, Data Everywhere.” *Economist* 394(8671): 1–18.
- Dickinson, Katherine, Richard West, Deborah Kogan, David Drury, Marlene Franks, Laura Schlichtmann, and Mary Vencill. 1988. *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*.

- Research Report No. 88–16. Washington, DC: National Commission for Employment Policy.
- Finn, D. 2008. *The British “Welfare Market”: Lessons from Contracting Out Welfare to Work Programmes in Australia and the Netherlands*. York, UK: Joseph Rowntree Foundation. <http://www.jrf.org.uk/sites/files/jrf/2306-welfare-unemployment-services.pdf> (accessed June 10, 2010).
- General Accounting Office (GAO). 1991. *Job Training Partnership Act: Racial and Gender Disparities in Services*. Report to the Chairman, Legislation and National Security Subcommittee, Committee on Government Operations, House of Representatives. HRD-91-148. Washington, DC: GAO.
- . 2002. *Improvements Needed in Performance Measures to Provide a More Accurate Picture of WIA’s Effectiveness*. Report to Congressional Requesters. GAO-02-275. Washington, DC: GAO.
- Heckman, James, Carolyn Heinrich, and Jeffrey Smith. 2002. “The Performance of Performance Standards.” *Journal of Human Resources* 37(4): 778–811.
- Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. “The Economics and Econometrics of Active Labor Market Programs.” In *Handbook of Labor Economics*, Vol. 3A, Orley Ashenfelter and David Card, eds. Amsterdam: North Holland, pp. 1865–2097.
- Heckman, James, and Jeffrey Smith. 2004. “The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program.” *Journal of Labor Economics* 22(4): 243–298.
- Heinrich, Carolyn J. 2004. “Improving Public-Sector Performance Management: One Step Forward, Two Steps Back?” *Public Finance and Management* 4(3): 317–351.
- . 2007. “False or Fitting Recognition? The Use of High Performance Bonuses in Motivating Organizational Achievements.” *Journal of Policy Analysis and Management* 26(2): 281–304.
- Heinrich, Carolyn J., and Gerald R. Marschke. 2010. “Incentives and Their Dynamics in Public Sector Performance Management Systems.” *Journal of Policy Analysis and Management* 29(1): 183–208.
- Klerman, Jacob Alex. 2005. “Measuring Performance.” In *High Performance Government: Structure, Leadership, Incentives*, Robert Klitgaard and Paul C. Light, eds. Santa Monica, CA: RAND Corporation, pp. 343–380.
- Koning, Pierre, and Carolyn J. Heinrich. 2010. “Cream Skimming, Parking, and Other Intended and Unintended Effects of Performance-Based Contracting in Social Welfare Services.” IZA Discussion Paper No. 4801. Bonn, Germany: IZA.
- Light, Paul C. 2005. *The Four Pillars of High Performance*. New York: McGraw Hill.

- Moynihan, Donald. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Orr, Larry, Howard Bloom, Stephen Bell, Winston Lin, George Cave, and Fred Doolittle. 1995. *The National JTPA Study: Impacts, Benefits, and Costs of Title IIA*. Bethesda, MD: Abt Associates.
- Propper, Carol, and Deborah Wilson. 2003. "The Use and Usefulness of Performance Measures in the Public Sector." *Oxford Review of Economic Policy* 19(2): 250–267.
- Radin, Beryl. 2000. "The Government Performance and Results Act: Square Pegs in Round Holes?" *Journal of Public Administration Research and Theory* 10(1): 11–35.
- Smith, Peter. 1995. "On the Unintended Consequences of Publishing Performance Data in the Public Sector." *International Journal of Public Administration* 18(2–3): 277–310.
- Struyven, L., and G. Steurs. 2005. "Design and Redesign of a Quasi-Market for the Reintegration of Jobseekers: Empirical Evidence from Australia and the Netherlands." *Journal of European Social Policy* 15(3): 211–229.

The Performance of Performance Standards

James J. Heckman
Carolyn J. Heinrich
Pascal Courty
Gerald Marschke
Jeffrey Smith
Editors

2011

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

The performance of performance standards / James J. Heckman . . . [et al.], editors.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-88099-292-3 (pbk. : alk. paper)

ISBN-10: 0-88099-292-1 (pbk. : alk. paper)

ISBN-13: 978-0-88099-294-7 (hardcover : alk. paper)

ISBN-10: 0-88099-294-8 (hardcover : alk. paper)

1. Government productivity. 2. Performance standards. 3. Civil service—Personnel management. I. Heckman, James J. (James Joseph)

JF1525.P67P476 2011

352.6'7—dc22

2011007877

© 2011

W.E. Upjohn Institute for Employment Research

300 S. Westnedge Avenue

Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the author. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by Alcorn Publication Design.

Index prepared by Diane Worden.

Printed in the United States of America.

Printed on recycled paper.