

RAND RESEARCH AREAS

CHILDREN AND FAMILIES
 EDUCATION AND THE ARTS
 ENERGY AND ENVIRONMENT
 HEALTH AND HEALTH CARE
 INFRASTRUCTURE AND TRANSPORTATION
 INTERNATIONAL AFFAIRS
 LAW AND BUSINESS
 NATIONAL SECURITY
 POPULATION AND AGING
 PUBLIC SAFETY
 SCIENCE AND TECHNOLOGY
 TERRORISM AND HOMELAND SECURITY

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of published, peer-reviewed documents.

Corporate Headquarters
 1776 Main Street
 P.O. Box 2138
 Santa Monica, California
 90407-2138
 TEL 310.393.0411
 FAX 310.393.4818

© RAND 2010

Are Performance-Based Accountability Systems Effective?

Evidence from Five Sectors

During the past two decades, performance-based accountability systems (PBASs), which link financial or other incentives to measured performance as a means of improving services, have gained popularity among policymakers. For example, the No Child Left Behind (NCLB) Act of 2001 (Pub. L. 107-110) combined explicit expectations for student performance with well-aligned tests to measure achievement, and it included strong consequences for schools that did not meet performance targets. In the transportation sector, cost plus time (A+B) contracting has become a popular means of streamlining and speeding up highway construction projects, while, in health care, there are more than 40 hospitals and more than 100 physician/medical group performance-based accountability programs (popularly dubbed pay-for-performance, or P4P) in the United States.

Although PBASs can vary widely across sectors, they share three main components: goals (i.e., one or more long-term outcomes to be achieved), incentives (i.e., rewards or sanctions to motivate changes in behavior to improve performance), and measures (formal mechanisms for monitoring service delivery or goal attainment).

But, while the use of PBASs has spread in the public sector, little is known about whether such programs are having the desired effect or how to design them to be as effective as possible. To address this gap, a RAND study examined several examples of PBASs, large and small, from a range of public service areas. The study focused on nine PBASs, drawn from five sectors: child care, education, health care, public health emergency preparedness (PHEP), and transportation.

The study suggests that PBASs represent a promising policy option for improving the quality of service delivery in many contexts. However, evidence of PBAS effectiveness is rare, and successful design requires careful atten-

Key findings:

- Under the right circumstances, a performance-based accountability system (PBAS) can improve the effectiveness and efficiency of services for the public; however, existing evidence for PBAS effectiveness is rare.
- PBASs are often created without consensus on key design issues.
- The selection of incentives and performance measures has proved challenging.
- Creating an effective PBAS requires careful attention to selecting an appropriate design, given the context in which it is to operate, and to monitoring, evaluating, and adjusting the system, as appropriate.

tion to the selection of incentives, performance measures, and implementation issues, as well as rigorous evaluation to monitor the program's effectiveness.

Methods

The PBASs studied are listed by sector in the table. The selection was guided by an interest in focusing on services in which the public sector has an important role, either in providing the services (e.g., education, transportation, PHEP) or in governance of the services (health care, child care).

The research approach included development of an analytic framework, a broad review of literature related to performance measurement and accountability in the private and public sectors, and an integrative workshop with experts and practitioners to examine various features of PBASs.

Cases Examined in the Study

Sector	PBASs
Child care	Quality rating and improvement systems (QRISs)
Education	NCLB P4P
Health care	Hospital and physician/medical group P4P programs, including quality report cards
PHEP	Centers for Disease Control and Prevention (CDC) PHEP cooperative agreement
Transportation	A+B highway construction contracting Corporate Average Fuel Economy (CAFE) standards Clean Air Act (CAA) ambient air pollution conformity requirements Transit subsidy allocation formulas

Key Findings

PBASs Can Be Effective in the Right Circumstances

Evidence on the effects of nine PBASs in five sectors shows that, under the right circumstances, a PBAS can be an effective strategy for improving the delivery of services to the public. According to a broad review of the literature and specific studies, certain optimal circumstances and design choices support an effective PBAS, as shown in the box.

Perhaps the most unambiguous example of an effective PBAS is A+B contracting for highway construction, in which contractors receive a financial bonus for completing road and construction projects within an accelerated time frame. Yet, A+B contracting presents, in many ways, a “best case” set of circumstances for an effective PBAS, including the widely shared goal of reducing construction time and a relatively unambiguous performance measure—days to

Elements That Contribute to an Effective PBAS

- A goal that is widely shared among all stakeholders
- Measures that are unambiguous and easy to observe
- Incentives that apply to individuals or organizations with control over the relevant inputs and processes
- Incentives that are meaningful to those being incentivized
- Few competing interests or requirements
- Adequate resources to design, implement, modify, and operate the PBAS

complete the project. Further, those held accountable under A+B contracting—construction firms—have near-complete control over the relevant inputs and processes involved in road construction, have the technical expertise to do the work, and know the health and safety standards against which their work will be judged. However, these ideal conditions are rarely fully realized, so it is difficult to design and implement PBASs that are uniformly effective.

Overall, the Study Found Limited Evidence About PBAS Effectiveness

In general, the study found limited evidence about the effectiveness of PBASs, which typically have not been subject to rigorous evaluation. The evidence that does exist leads to somewhat different conclusions by sector. For example, in education, it was clear that NCLB and other high-stakes testing programs with public reporting and other incentives at the school level have led to changes in teacher behavior; however, teachers seem to have responded narrowly in ways that improve measured outputs while paying less attention to long-term outcomes or goals. In health care, relatively small financial incentives (frequently combined with public reporting) have had some modest effects in improving the quality of care delivered. In transportation, large financial incentives have led to creative solutions, as well as lobbying to influence the demands of PBAS regulation. It is too soon to judge the effectiveness of the PBASs we studied in child care and PHEP.

PBASs Were Often Created Without Consensus on Key Design Issues

The study found that a strong knowledge base about the drivers of performance in the sector helped create consensus about who should be held accountable for what. In the sectors we examined, however, there were often differences of opinion about the desirability and general contours of PBASs. It appears that PBASs were often created in spite of a lack of consensus about key issues.

To be measured, performance must be defined precisely. However, the study found only a general understanding of performance in several of the cases examined. In child care, for example, there was general agreement about the broad goal (i.e., to improve the quality of care) but little agreement on specifics, such as which outcomes matter most (e.g., “kindergarten readiness,” learning to regulate emotions, ability to follow instructions). The story was similar in education, with agreement on broad goals (e.g., to produce high-school graduates with high levels of achievement, advanced skills) but not on the specific performance elements that should be assessed.

Decisions must be made in terms of who should be held accountable. In several cases, conflicts arose regarding

the most appropriate people to target for behavioral change through the PBAS. For example, in health care, should the focus be on the physician, the practice site, the larger medical group, or an integrated delivery system of physician groups and hospitals? Generally, service providers prefer to be held accountable only for those aspects of service production over which they have clear and direct control. In PHEP, for instance, health departments caution against PBASs that would hold them responsible for maintaining security and argue that they should be held accountable only for such activities as building partnerships and coordinating with law-enforcement and security agencies.

Selection of Incentive Structures Has Proved Challenging

Context appeared to have a large effect on the incentive structures that PBAS designers chose. In our sample, when participation in a PBAS was voluntary, designers of PBASs typically used rewards rather than sanctions (e.g., child-care QRISs, A+B contracting), while sanctions were more common when designers worked in a regulatory setting (e.g., NCLB).

The review showed that a PBAS might not stimulate a significant provider response if the incentives are not large enough. For example, in many health-care P4P programs, the potential financial rewards represent a very small percentage of overall physician pay and thus may not garner much attention. Across the PBASs studied, it was unclear how well the magnitude of rewards was correlated with the benefits of the changes that the PBAS designers sought to induce or with the effort required of service providers.

The structure of incentives can give rise to unanticipated and undesired consequences. To cite one example, in NCLB, attaching public reporting and other incentives to test scores has led, in some cases, to unintended behavior changes (i.e., “teaching to the test”) that might be considered undesirable.

Design of Performance Measures Required Trade-Offs Among Competing Priorities

PBAS designers made trade-offs among a number of competing factors when they selected and structured measures, including, among other things, the feasibility, availability, and cost of measures; the context within which a PBAS operates; and the degree and locus of control.

PBAS designers typically paid close attention to costs. Designers typically avoided measures that would be very expensive to collect (unless the measures were already captured for some other purpose). In health care, for example, the most detailed and complete outcome measures would require costly manual review and data extraction from numerous medical charts; accordingly, many health-care PBASs instead used less expensive surrogates (e.g., measures

of inputs or outputs). PBAS designers also often limited the number of measures in order to monitor and evaluate costs.

To the extent possible, PBASs sought to incorporate existing measures. In allocating transit funding, PBASs made use of statistics that were already reported to the Federal Transit Administration’s National Transit Database. In health care, many PBASs used standard quality measures developed by the National Committee for Quality Assurance.

PBAS designers often relied on the best available measure, whether or not it was the best measure. The outcomes that a PBAS seeks to affect often unfold far into the future. As a result, many PBASs measure short-term outputs rather than desired outcomes that are realized only after many years. For example, child care presumably influences the child’s future character development, educational attainment, and skill sets. But, even if a child-care PBAS seeks to improve these outcomes, it cannot measure, during an accountability cycle, how the actions of a child-care provider will ultimately affect the children the provider currently serves.

Recommendations for PBAS Developers

The study offered a range of recommendations for PBAS sponsors, designers, and other stakeholders regarding PBAS design, incentives, performance measurement, implementation, and evaluation.

Design

Consider the factors that might hinder or support PBAS effectiveness to see whether conditions support system development and use. A PBAS is not always the best option for improving performance. If a large share of the factors that support effective implementation do not exist, decisionmakers may wish to consider alternative policy options or think about ways to influence the context to create more-positive implementation conditions for a PBAS.

Be sensitive to the context for implementation. PBAS designers should attempt to understand the drivers of performance in the service activity they are seeking to improve. Such knowledge can support the development of consensus about who should be held accountable for what.

Incentives and Performance Measures

Select the right unit of accountability. Incentives and performance measures should focus on appropriate units of accountability (e.g., individual, department, organization). In some cases, it may be useful to apply incentives and performance measures at different functional levels within a service activity (e.g., in education, set up different performance measures and incentives for school districts, school principals, and teachers).

Make the rewards or penalties big enough to matter without exceeding the value of improved performance.

Many options for incentives are available, including cash, promotions, status, recognition, increased autonomy, and access to training or other investment resources. The goal is to select options that will best influence behavior without undermining intrinsic service motivation. The size of the incentive should be greater than the effort required of the service provider to improve on the performance measure but should not exceed the value obtained from improved provider behavior.

Focus on performance measures that matter. Performance measures determine how service providers focus their efforts. To the extent possible, therefore, it makes sense to include those measures believed to have the greatest effect on the broader goals of interest.

Create measures that people can influence. Individuals or organizations should not be held accountable for things they cannot control. PBAS designers typically have multiple relevant options for measures, including output measures that account for relevant social, physical, or demographic characteristics of the population served; measures that are based on inputs, structure, or processes rather than outputs or outcomes; and measures of relative improvement instead of absolute performance.

Implementation and Evaluation

Implement the program in stages, and modify as needed.

To obtain the best results for the long term, it is important to develop a plan for monitoring the PBAS, identifying shortcomings that may be limiting PBAS effectiveness or leading to unintended consequences, and modifying aspects of the program (e.g., incentives, measures, units of accountability) as needed. Pilot-testing might be used to assess measures and other design features.

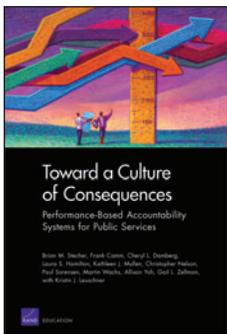
Integrate the PBAS with existing performance databases and accounting and personnel systems. It is important to think through all the ways in which the PBAS will interact with existing infrastructure—e.g., performance databases, accounting systems, personnel systems. These considerations may suggest changes in the PBAS design or highlight ways in which the existing infrastructure needs to be modified.

Engage service providers and, to the extent possible, secure their support. The use of credible, fair, and actionable measures can help garner the support of service providers whose performance is to be measured. Service providers might be asked to participate in the process of developing the measures and incentives. Regular communication is also key.

Evaluation is critical. Only through careful monitoring and evaluation can decisionmakers detect problems and take steps to improve PBAS functioning over time. The evaluation should be structured according to the PBAS's stage of development. For example, when a system is first developed, it may be most helpful to evaluate implementation activities (e.g., whether appropriate mechanisms for capturing and reporting performance measures have been developed). As the system matures, the focus could shift to evaluating the performance measures' and incentive structure's effects (e.g., observed provider behavior and service outputs).

Conclusion

The study suggests that PBASs represent a promising policy option for improving the quality of service-delivery activities in many contexts, and the evidence supports continued experimentation with and adoption of this approach. However, the prospects of effectiveness for a PBAS are highly dependent on the context in which it is to be implemented. Thus, careful attention should be paid to selecting an appropriate design for the PBAS and to monitoring, evaluating, and adjusting the system, as appropriate. ■



This research brief describes work done for RAND Education documented in *Toward a Culture of Consequences: Performance-Based Accountability Systems for Public Services*, by Brian M. Stecher, Frank Camm, Cheryl L. Damberg, Laura S. Hamilton, Kathleen J. Mullen, Christopher Nelson, Paul Sorensen, Martin Wachs, Allison Yoh, and Gail L. Zellman, with Kristin J. Leuschner, MG-1019 (available at <http://www.rand.org/pubs/monographs/MG1019/>), 2010, 272 pp., \$28, ISBN: 978-0-8330-5015-1; and *Toward a Culture of Consequences: Performance-Based Accountability Systems for Public Services—Executive Summary*, by Brian M. Stecher, Frank Camm, Cheryl L. Damberg, Laura S. Hamilton, Kathleen J. Mullen, Christopher Nelson, Paul Sorensen, Martin Wachs, Allison Yoh, and Gail L. Zellman, with Kristin J. Leuschner, MG-1019/1 (available at <http://www.rand.org/pubs/monographs/MG1019.1/>), 2010, 50 pp., \$15, ISBN: 978-0-8330-5016-8. This research brief was written by Kristin J. Leuschner. The RAND Corporation is a nonprofit research institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

RAND Offices

Santa Monica, CA • Washington, DC • Pittsburgh, PA • New Orleans, LA/Jackson, MS • Boston, MA • Doha, QA • Cambridge, UK • Brussels, BE



EDUCATION

THE ARTS
CHILD POLICY
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND
HOMELAND SECURITY
TRANSPORTATION AND
INFRASTRUCTURE
WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of individual published, peer-reviewed documents or of a body of published work.

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Education](#)

View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).