

NOTES ON THE APPAM-MOSCOW CONFERENCE

IMPROVING THE QUALITY OF PUBLIC SERVICES: A MULTINATIONAL CONFERENCE ON PUBLIC MANAGEMENT

Douglas J. Besharov, Alexey Barabashev, Karen Baehler, and
Jacob Alex Klerman

In June 2011, the Association for Public Policy Analysis and Management, in association with Moscow's National Research University—Higher School of Economics, and the University of Maryland School of Public Policy, held an international research conference in Moscow on “Improving the Quality of Public Services and Public Management.” Attendees came from more than 22 nations, including Australia, Austria, Canada, China, Finland, Germany, Israel, Italy, Lithuania, Luxembourg, Mexico, Moldova, Philippines, Republic of Korea, Russia, Slovakia, South Africa, Spain, Turkey, the United Arab Emirates, the United Kingdom, and the United States.

Ninety-two papers were presented on public management and its ability to improve the quality of public services. The most common topics were performance measurement and management, citizen participation, and transparency and anti-corruption efforts in post-socialist countries. Other topics included e-government, inter-agency collaboration, program evaluation, public-private partnerships, privatization, performance contracting, leadership, administrative modernization, networks, and the continuing relevance of the New Public Management paradigm. Many of the papers also explored the role of civil society in post-socialist countries. All of the conference papers can be found at: http://umdcipe.org/conferences/Moscow/moscow_papers.html. Two of the papers are published in this volume (Evans & Campos, 2013; Sanger, 2013).

Most striking were the similarities in the problems faced by public managers around the world and the parallel approaches many have taken in response. Although coming from many different cultures and political situations, the presenters shared many points of agreement. Nowhere was this more evident than in the papers on the theory and practice of performance measurement. We do not have space to summarize all the papers presented, but four themes bear emphasis: (1) the need for performance measurement to make greater use of the tools of program evaluation; (2) how performance measurement could improve public management; (3) the continuing challenge of connecting performance goals to outcomes; and (4) the application of performance measurement techniques to incentive-based grants and contracts.

PERFORMANCE MEASUREMENT AND PROGRAM EVALUATION

Many conference papers explicitly explored the interrelation between program evaluation and performance measurement; others noted that learning what works (evaluation) is often essential to successful public management. Papers covered performance measurement for human services and welfare programs; creating more objective, systematic, and uniform performance measurement systems; using performance measurement for planning, managing, and budgeting; using surveys to measure performance; and measuring the success (or failure) of incentive-based contracts.

Journal of Policy Analysis and Management, Vol. 32, No. 1, 204–210 (2013)

© 2012 by the Association for Public Policy Analysis and Management

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam

Supporting Information is available in the online issue at wileyonlinelibrary.com.

DOI:10.1002/pam.21672

In his plenary talk, “The Intersection of Performance Measurement and Program Evaluation: Searching for the Counterfactual,” Douglas Besharov of the University of Maryland argued that, besides measuring program outputs, performance measures should measure outcomes through a counterfactual, i.e., what would have happened without the program.¹

Besharov noted that, although conventional management focuses on inputs and outputs, outputs do not necessarily lead to the desired outcomes. Hence, focusing on outputs is insufficient, and managers should also measure and manage to outcomes. His parallel argument (that outcomes do not always lead to longer-term impacts) would imply that we should manage to impacts rather than to outcomes. But, Besharov argued, impacts usually occur too late to be useful for real-world and real-time performance management. Instead, he advocated using plausible logic models—ideally supported by high-quality scientific theory or evaluation evidence—to choose outcomes for which the link to subsequent impacts is well supported.

Besharov’s proposal rests on the ability of public managers to identify and measure counterfactuals accurately. He suggested that performance measurement draw ideas from the field of program evaluation, including pre- and post-comparison, difference-in-differences, regression-discontinuity, and forms of randomization.

Integrating program evaluation ideas with performance measurement and management will require sharp divergence from current practice, at least in the United States, according to Maureen Pirog of Indiana University and the University of Washington. In her plenary talk, Pirog noted an absence of counterfactual thinking in most performance-incentive programs (2011). As a result, performance bonuses often do not go to the most effective programs, and misplaced performance incentives may end up hurting the worst-off clients.

Jacob Klerman of Abt Associates pursued this argument in his plenary talk, “COI for Social Programs.” Although an approximate counterfactual may be sufficient for performance measurement and management, for program evaluation, an exact counterfactual is needed. Random-assignment is the most reliable way to generate such a counterfactual. The challenge, then, is to generate counterfactuals and measure them—within regular program operations.

According to Klerman, a major impediment to evaluations (especially evaluations to identify ways of incrementally improving existing programs) is the cost of data collection for the very large samples required for such evaluations to detect the likely small changes induced by an implementation adjustment (i.e., smaller than the program and no program comparisons of conventional up-down program evaluations). He noted, however, that when both the treatment and the control and comparison groups are in the program, a performance measurement system can solve the data-collection problem for the outcomes to which the program manages. Such near-zero incremental costs for data collection would make it feasible to mount evaluations large enough to detect the likely small impacts of incremental changes. Klerman concluded with suggestions for how a program might embed such site-level mini-experiments (and evaluations) in ongoing program operations.

Randomized experiments may not always be possible. “Workforce Development Program Performance over the Business Cycle” involved the use of a matched comparison group to identify the counterfactual. Kevin Hollenbeck of W.E. Upjohn Institute for Employment Research, and David Pavelchek of the Washington State Workforce Training and Education Coordinating Board (WTECB) used 10 years

¹ Besharov’s analysis begins with a logic model in which programs use inputs (e.g., staff and facilities) to provide outputs (e.g., hours of service provided), that in turn lead to outcomes (i.e., short-term effects on clients) and impacts (i.e., long-term effects on clients).

of administrative data from WTECB to explore the cyclicity of job placement outcomes for participants in various state-supported work-training programs between 1998 and 2007.

Hollenbeck and Pavelchek found the impact of training programs (measured by placement rates for participants compared with a matched comparison group) to be mostly procyclical during the 2000s. As expected, longer duration programs that focused on building human capital produced more stable outcomes than shorter duration programs that focused on immediate job placement. Besides contributing to the literature on job training programs, their work demonstrates the potential of quasi-experimental methods in performance measurement.

In “Making Services Work: Indicators, Assessments, and Benchmarks of the Quality and Governance of Public Service Delivery in the Human Development Sectors,” Ariel Fiszbein, Dena Ringold, and Halsey Rogers drew on their experience at the World Bank to develop a typology of measurements for human services. Like many other presenters, they used a logic model to map how government policies lead to government performance, which in turn leads to the quality of service delivery, and finally to human development outcomes.

Measures should include government policies, government functioning at the organizational level, the quality of service delivery, and human development outcomes at the client level, according to Fiszbein, Ringold, and Halsey. For each, they presented examples of successful measurements, many of which could be directly usable by others and the remainder could serve as models for the development of similar measurement systems.

In her plenary talk, “System Mapping as a Tool of Organizational Learning and Change,” Karen Baehler of American University echoed other speakers’ endorsements of the logic model, but urged greater attention to understanding the larger systems in which logic models reside. In order to explain why a program or agency is performing well or poorly, Baehler argued, scholars and practitioners need to know a great deal about the clients being served, their problems, the factors capable of mitigating program impacts, and critical features of the program’s setting.

LINKING PERFORMANCE MEASUREMENT TO OUTCOMES

The challenge of connecting performance goals to outcomes was another common theme. Papers covered matching performance measurement and management reforms to institutional realities, social service oversight and regulation in Israel, whether or not measuring performance leads to better performance, and the relationship between management and outcomes (including NYC’s welfare-to-work program and the Turkish health care system).

In “Matching Reforms to Institutional Realities: A Framework for Assessing Social Service Delivery Reform Strategies in Developing Countries,” for example, Ariel Fiszbein and Yasuhiko Matsuda, both of the World Bank, developed a catalog of institutional options that can be used by disinterested outsiders (e.g., international donors) to manage performance and thereby improve the delivery of social services. These include inter-governmental transfers, enforcement of hard budget constraints for local governments, clear definitions of responsibilities, and monitoring and evaluation.

Beyond their catalog, they presented a theory to identify the country-specific conditions that would optimize each institutional option—with the appropriateness of different strategies varying with the degree of local government accountability (usually electoral), central government incentives for local governments, and central government institutions to control local governments. The paper is informed by rich examples drawn from the World Bank’s work.

Bryna Sanger (2013) of The New School was less sanguine. In “Does Measuring Performance Lead to Better Performance?,” she reported on a survey of 198 jurisdictions and a review of their public documents that she used to rank them in terms of their use of best measurement practices. From these, she chose for more intensive study 24 jurisdictions that had adopted the best practices.

Sanger asserted that true performance management requires frequent measurement, monthly in many circumstances. Only eight of her 24 jurisdictions, however, claimed to look at data monthly; and only three looked at data weekly. Moreover, like many of the other participants, she is a proponent of performance measurement and performance management systems that focus on outcomes (or intermediate impacts), rather than inputs or longer-term outcomes. To address the counterfactual problem, she noted that many systems use benchmarking, i.e., they compare their performance against the performance of similar organizations in other jurisdictions performing the same tasks or across units within their own jurisdiction. Outliers on both extremes can be used to identify which units might be the subject of emulation or remediation.

Even in this top-rated group, Sanger found no evidence that measuring performance led to better performance, and she expressed a fear that even ideal performance measurement systems might not lead to strong performance management. Worse, she found a lack of sustained interest in performance management and decreasing investment in it because of the budget cutbacks resulting from the Great Recession.

One interpretation of Sanger’s results is that systems that do not attach consequences to performance measures are unlikely to influence outcomes. The converse was the case in a reform of the Turkish health system’s method of compensating health professionals, where, in addition to receiving salaries, they also received potentially large performance bonuses each month (several times base pay). According to Gulbiye Yasar of Ankara University and Pinar Guven-Uslu of the University of East Anglia, in “Performance-Based Supplementary Payment System in Ministry of Health Practices in Turkey,” the reforms operated as intended, increasing hours worked in the ministry’s facilities (less moonlighting and private practice) and increasing the number of patients served.

Here, the link between performance and outcomes seems clear: Performance measures quantify professional work effort, performance management converts these measurements into higher pay, and the promise of higher pay leads to more hours and better client outcomes (assuming that medical care is effective). There is, however, some evidence of unintended side effects: more competition between professionals, less team work, greater reluctance to admit complex patients to the hospital, and less effort toward activities that do not count.

PERFORMANCE MEASUREMENT AND PERFORMANCE-BASED FUNDING

Several conference papers applied concepts of performance measurement to performance-based grants and contracts. Papers covered holding nonprofits accountable for results, the efficiency of contracting out local public services (including ambulance services and shelter services for children of illegal immigrants), performance-based contracting in NYC’s welfare-to-work program, contracting out energy provision in Europe, and nonprofit cooperation in Russia.

In “Holding Non-Profits Accountable: Frameworks that Distinguish between Looking Good and Doing Good,” Christina Standerfer of the University of Arkansas and Joseph Schaefer, an independent nonprofit consultant, called on charity watchdog groups to join the movement for more robust nonprofit accountability by including outcome-based indicators in their rating reports. Like many other speakers, they

emphasized that input- and output-based measures do not show whether services improve the well-being of individuals, communities, or the environment.

Standerfer and Schaefer used a logic model to identify performance measures for nonprofit organizations. After describing and critiquing four performance and accountability frameworks from the literature, they presented their own hybrid, which specifies outcomes tied to specific, observable, and verifiable indicators—tailored to the accountability demands of different stakeholder groups. In contrast, both the standard government grant reporting template and the information fed into monitoring and rating Web sites such as Charity Navigator focus on what the authors call “dollars and widgets,” i.e., financial balance sheets (inputs) and service units (outputs), rather than outcomes for people.

In “The Efficiency of Contracting out Local Public Services,” Beata Merickova of Matej Bel University, Slovakia; Juraj Nemeč of Masaryk University, Czech Republic; and Zuzana Vozarova also of Matej Bel University compared the performance of publicly provided services with contracted services in a sample of Czech and Slovak municipalities between 2000 and 2009.

Measured through a combination of cost indicators (per resident and per unit of service) and quality indicators (based on citizen satisfaction surveys), the relative efficiency of publicly provided vs. contracted services was mixed. Across six groupings of municipalities based on size and five categories of services, internalized delivery of services performed better than contracting in 60 percent of cases, with the major exception being the public lighting service.

The authors concluded that contracting out in transitional countries seemed less successful than theory would predict, perhaps because of the persisting market and social conditions in many countries transitioning from socialism to capitalism. For example, some areas had too few private suppliers, leading to very low numbers of bids being submitted to municipalities in some tender processes; and too many municipalities did not comply with national procurement laws, instead purchasing services directly from a preferred supplier.

The experience with public–private partnerships around the world has generated great interest in the best ways to measure their success and failure. In “The Design and Implementation of Public-Private Partnerships in the U.K.’s Social Sector,” Tahir Nisar of the University of Southampton described the UK’s efforts to incorporate relevant performance measures into its private finance initiative (PFI), and then assessed the results in two case studies: (1) a design-build-finance-operate project for a community school and (2) a contract to maintain and operate a nursing home for the elderly.

Nisar found that both projects experienced implementation challenges that the PFI screening criteria failed to anticipate. To reduce the probability of such problems in the future, Nisar recommended phased reviews of implementation progress and an expansion of screening criteria to include project quality standards and a review of project control mechanisms.

In “Welfare-to-Work Performance-Based Contracts in New York City: Lessons Learned,” Swati Desai of Columbia University, and Lisa Garabedian and Karl Snyder, both of the New York City Human Resources Administration reviewed the city’s experience with contracting for welfare-to-work services across three successive sets of contracts. The contracts were vigorously performance based, and paid close attention to how much the city paid for specific outcomes in an attempt to guide contractor effort toward desired outcomes.

From this experience, they identified four lessons: (1) the design of contract payment milestones is critical; (2) technology and performance measurement systems are essential to managing contracts successfully; (3) individual contractors do not behave identically and contract design and management decisions must account for

varying contractor approaches; and (4) officials should maximize contract flexibility and learn from past performance.

The conference's Moscow venue offered participants an unparalleled opportunity to hear from noted Russian scholars and practitioners about the country's public service reform agenda, accomplishments, and continuing challenges. For example, two plenary talks, by Alexey Barabashev "Civil Servants Effectiveness Evaluation: Russian Experience" and Andrey Klimenko "Public Administration Development in Russia: Major Results and Perspectives," both of the National Research University—Higher School of Economics, examined the complications of performance measurement and program evaluation in a country with highly centralized power, widespread mistrust of government, limited access to information and services, commonplace corruption, and excessive governmental regulation.

Both Barabashev and Klimenko voiced confidence that a more effective and efficient bureaucracy would follow from a clearer distribution of responsibilities, improved access to government information, more quality control by citizens, recruitment of highly qualified personnel (and possible outsourcing of certain functions), streamlining the service-delivery process, and independent evaluation of the quality of government services.

One hopes conferences like this will further this goal worldwide, as well as in Russia.

DOUGLAS J. BESHAROV is a Professor at the School of Public Policy, University of Maryland, 2101 Van Munching Hall, College Park, MD 20742.

ALEXEY BARABASHEV is a Professor at the School of Public Administration and Municipal Management, National Research University—Higher School of Economics, 20 Myasnitskaya Str., Moscow, 101000 Russia.

KAREN BAEHLER is a Scholar in Residence at the School of Public Affairs, American University, 4400 Massachusetts Avenue, NW, Washington, DC 20016.

JACOB ALEX KLERMAN is a Principal Associate of Social & Economic Policy at Abt Associates, Inc., 55 Wheeler Street, Cambridge, MA 02138.

REFERENCES

- Baehler, K. (2011, June). System mapping as a tool of organizational learning and change. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Barabashev, A. (2011, June). Civil servants effectiveness evaluation: Russian experience. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Besharov, D. (2011, June). The intersection of performance measurement and program evaluation: Searching for the counterfactual. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Desai, S., Garabedian, L., & Snyder, K. (2011, June). Welfare-to-work performance-based contracts in New York City: Lessons learned. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management*, 32, 179–191.
- Fiszbein, A., & Matsuda, Y. (2011, June). Matching reforms to institutional realities: A framework for assessing social service delivery reform strategies in developing countries. Paper

- presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Fiszbein, A., Ringold, D., & Rogers, H. (2011, June). Making services work: Indicators, assessments, and benchmarks of the quality and governance of public service delivery in the human development sectors. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Hollenbeck, K., & Pavelchek, D. (2011, June). Workforce development program performance over the business cycle. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Klerman, J. (2011, June). CQI for social programs. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Klimenko, A. (2011, June). Public administration development in Russia: Major results and perspectives. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Merickova, B., Nemeč, J., & Vozarova, Z. (2011, June). The efficiency of contracting out local public services. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Nisar, T. (2011, June). The design and implementation of public-private partnerships in the U.K.'s social sector. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Pirog, M. (2011, June). The role of policy analysis research in U.S. public policy. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Sanger, M. B. (2013). Does measuring performance lead to better performance? *Journal of Policy Analysis and Management*, 32, 192–209.
- Standerfer, C., & Schaefer, J. (2011, June). Holding non-profits accountable: Frameworks that distinguish between looking good and doing good. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.
- Yasar, G., & Guven-Uslu, P. (2011, June). Performance-based supplementary payment system in Ministry of Health practices in Turkey. Paper presented at Improving the Quality of Public Services: A Multinational Conference on Public Management, Moscow, Russia.