The World Bank is often called on by governments to assist them with the process of reforming their performance management systems and measurement approaches. Drawing on practical experience from World Bank operations, this issue of Governance Notes presents methods for assessing the effects of results-based management (RBM) reforms on outcomes and processes.

Three complementary methods are described: impact evaluations and nonexperimental methods, in-depth interviews with government officials, and surveys with civil servants. The strengths and weaknesses of each method are discussed, and practical guidance for application is provided. Given the central role processes play in reform initiatives, the authors conclude that combining quantitative and qualitative analysis will allow a better examination of both process and impact.

OBJECTIVES AND TOOLS
Results-based management encompasses a wide variety of approaches, techniques, and tools used by public sector agencies to monitor, evaluate, and produce results (Dorotinsky and Watkins 2011, 1). RBM reforms can bring about improvements in several ways, such as aligning different agencies along common objectives and tracking progress against agreed-to goals, or creating an environment in which organizations learn from each other and improve program design.

The principal objectives for employing RBM approaches include gaining a better understanding of budget expenditure outcomes, incentivizing agencies to link outputs to impacts, and prompting behavior change in public sector agencies to deliver better services to citizens. Often, RBM systems are implemented against the backdrop of declining fiscal conditions, low public sector efficiency, and rising expectations on public service delivery (Kettl 2000, 5). RBM reforms are frequently introduced alongside other reforms (e.g., integrated financial management systems, medium-term expenditure frameworks, and human resources reforms), which can pose additional challenges to attributing observed impact.

Depending on the objectives being pursued, different RBM tools address different linkages and accountability arrangements among various actors within the public sector. Among the most commonly adopted tools are performance-related pay, strategic planning, benchmarking, total quality management, and performance-based budgeting (Vifuela and Zoratto 2015, 2). Many of these are mutually reinforcing, creating a network of supporting arrangements that can be challenging to untangle when measuring impact.

Measuring outcomes of reforms and providing evidence on what works, when, and how are imperative for enhancing public sector performance. This note compiles practical information on feasible ways to assess the impact of RBM reforms in World Bank operations.

WORLD BANK INVOLVEMENT

Globally, Latin America and the Caribbean stands out as a Region in which World Bank financed projects have helped generate valuable knowledge and practical experience on RBM systems. Selected projects in the Region with significant RBM components include the Brazil Minas Gerais Development Partnership Series (2006–2014) and Manaus Service Delivery and Fiscal Management Development Policy Loan (2014–2016).

In Brasil, several states and municipalities at the subnational level applied RBM approaches and incorporated reform elements into World Bank operations. During the preparation and execution of many of these projects, valuable knowledge on the implementation of RBM approaches was generated that could inform potential follow-up operations with similar objectives.

**METHODS FOR MEASURING IMPACTS**

Management reforms involve multidimensional institutional changes that evolve over time. Using one single method would draw only a partial view of effects. This note describes three complementary methods to assess the impact of RBM reforms. They have different strengths and limitations. A combination of methods is preferred to yield comprehensive results.

*Quantitative Analysis: Impact Evaluations and Nonexperimental Methods*

Impact evaluations (randomized control trials) at the individual or team level in a specific sector (e.g., at the student or school level) are able to establish a causal link between specific RBM reforms and observed results. For the education sector, examples include Bruns et al. (2012), which explores how performance-related pay influences the behaviors of teachers and ultimately affects students’ proficiency in Brazilian states, and Barrera-Osorio and Raju (2015), which evaluates a teacher performance pay program in Pakistan. For the security sector, Assis (2012) analyzes the effect of an RBM reform on public security outcomes in Minas Gerais. Unfortunately, impact evaluations are not always feasible because of limited data and resource availability, and ideally they should be embedded in World Bank operations from the start to avoid these obstacles.

A second, and less robust option, is to contrast current with pre-intervention outcomes such as child mortality, student proficiency, and homicide rates at the aggregate level (i.e., not per hospital, school, or police department). For example, Viviuela and Zoratto (2015) analyzed the impact of RBM reforms in Brazilian states by comparing outcomes with other states that had not implemented RBM systems. A truncated time-series model was applied to evaluate if discontinuities were statistically significant.

The internal validity of the analysis might be weak, however, since RBM reforms are usually implemented in a system- or government-wide fashion and not designed as randomized control trials with clear treatment and control groups. Furthermore, the effects of external factors, such as a change in the socioeconomic composition of the population, might be difficult to untangle from reform effects.

**IMPACT EVALUATIONS AND OTHER QUANTITATIVE METHODS**

| COST: | low/medium/high (high for impact evaluations) |
| STAFF HOURS: | low/medium/high |
| STRENGTHS: | large and representative samples, clear methods for inferring causality |
| WEAKNESSES: | Internal validity threats (e.g., omitted variables and selection bias), external validity in the case of impact evaluations |
| TOOLS: | statistical software (e.g., Stata and R) |
| KEY STEPS: | data collection (baseline and end line), impact evaluation design, data analysis and interpretation |

*Qualitative Analysis: In-Depth Interviews*

Quantitative analysis is the preferred method for measuring levels and changes in outcomes and for drawing inferences from observed statistical relations. It is less practical, however, in understanding process—the mechanisms by which a particular intervention instigates a series of events that ultimately result in the observed impact (Rao and Woolcock 2003, 167). To overcome this limitation, it is useful to complement the quantitative analysis with qualitative information to shed light on the chain of causality between the actions taken by the public sector and the observed changes in outcome indicators.

In-depth interviews can offer insights on what happened during the implementation process and help determine the extent to which “design failure” or “implementation failure” affected the achievement of intended results. Opening the black box of the implementation process is especially important for the evaluation of RBM reforms since they are centered on process issues.

---

1. These methods are analyzed in the Implementation Completion and Results Report for the Brasil Minas Gerais Development Partnership Series (World Bank 2015, annex X).
In-depth interviews include both individual interviews (e.g., one on one) and group interviews (e.g., focus groups). Both types provide an opportunity for collecting rich and detailed hypothesis about how individuals or smaller groups experience and understand the implementation of RBM reforms. For example, interviews were conducted to evaluate the process of establishing a central delivery unit to enhance results-based practices in Romania (P147482). Interviews provide more detail and depth than the quantitative analysis, and the standard survey and can be tailored specifically to the knowledge and position of the interviewee.

Format can range from less to more structured, depending on the specific research interest and available resources. More-structured interviews ensure that interviewees are asked the same questions to compare responses. This format increases generalizability and can serve to test specific hypotheses. Less-structured interviews, on the other hand, are most suitable for early stages of research because they allow interviewees to concentrate on what they think is most relevant to the question. The unstructured interview is generally very time consuming. Its use is advised where little is known about the topic of interest.

Focus groups share common features with less-structured interviews. They are comparatively easy to conduct and relatively inexpensive. A facilitator guides a group discussion on a topic. This approach can serve to help understand collective views on reforms and draw a hypothesis about the mechanism through which individual and collective behavior is affected. The task team can connect and monitor the discussion by video link to allow for a neutral discussion.

In-depth interviews can take on an explorative character during the early stage of research when little is known about the topic of interest. However, they can also be applied at a later stage to complement the findings of a quantitative analysis.

---

**IN-DEPTH INTERVIEWS**

**COST:** low/medium/high

**STAFF HOURS:** low/medium/high

**STRENGTHS:** flexibility, greater detail and depth

**WEAKNESSES:** time intensive, small number of observations, anecdotal evidence

**TOOLS:** interviewers, focus group facilitators, structured questionnaires

**KEY STEPS:** decide on type of in-depth interview; prepare questionnaire and conduct interviews; analyze and interpret findings

---

**Surveys**

Surveys can have either a more quantitative or qualitative orientation, depending on the specific research interest. In general, it is useful to follow a quantitative approach when they are administered to larger populations. In this case, well-structured and close-ended survey questions are used to facilitate the generation of numerical data that can be transformed into usable statistics. However, even predominantly quantitative surveys can have a qualitative element to them. For example, an open-ended question is often introduced to gather a better sense of the data as well as quotes to personalize the numerical data collected.

The online survey is the most inexpensive among the methods and tools of measurement. Technological advances make it possible to design, conduct, and analyze online surveys at a very low cost. Other advantages include (i) rapid deployment and return times since data on responses are available immediately; (ii) flexibility in design since complex surveys can be programmed easily, even using intricate skip patterns; and (iii) convenience for respondents, since they can answer questions on their own schedules. On the downside, online surveys may offer low response rates, especially if the target population has limited or no access to the internet (e.g., the elderly and people residing in remote areas). Because of these limitations, it’s best to use online surveys as a complementary method to other forms of measurement.

For example, during the evaluation of the Braşil Minas Gerais Development Partnership Series (2006–2014), an
online survey was conducted complementing qualitative face-to-face interviews and a quantitative analysis of public health, education, and security sector data. During project supervision missions, the task team heard enthusiastic accounts of a "change in culture" or "behavior change" resulting from the implementation of RBM reforms in the public administration of the Brazilian state of Minas Gerais.

This anecdotal evidence prompted a perception survey among civil servants to capture changes in behavior and motivation that might have resulted from the introduction and expansion of the RBM system in previous years. A link to the online survey was sent by email to the 17,000 public servants based in the administrative center of Minas Gerais; about 2,000 people completed the survey.

Despite certain limitations, such as the low response rate and reliance on answers potentially subject to recall bias, the findings of the survey provided valuable insights on the perception of the reforms by public employees. For instance, a large proportion of respondents believe the government and its staff had moved in the direction of working for results, instead of working to "fulfill processes," and the vast majority stated that the introduction of performance agreements improved their understanding of their role and goals at work, which in turn boosted their intrinsic motivation.

**CONCLUSIONS**

Results-based management reforms can take various forms and affect various sectors across public administration. Deciding on adequate methods of measuring reform outcomes depends on the specific research interest. If an evaluation’s objective is measuring a reform’s impact rather than understanding the processes by which the intervention worked, quantitative analysis (e.g., impact evaluation) may be the method of choice. However, given the central role that processes play in RBM reforms, it is advisable to complement the quantitative analysis with qualitative measures such as surveys, focus groups, and interviews to examine process as well as impact. A combined approach will help distinguish between design and implementation failures.

To accurately identify reform outcomes and the mechanisms through which they are obtained, apply a mixed-methods approach that captures the combined strengths of a set of quantitative and qualitative methods. It brings together “quantitative approaches that permit estimates of magnitude and distribution of effects, generalization and tests of statistical differences with qualitative approaches that permit in-depth description, analysis of processes and patterns of social interaction” (Woolcock, Rao, and Bamberger 2010, 11). Applying a well-designed mixed methods approach will provide the flexibility needed to fill information gaps, allow the use of triangulation to strengthen validity, and provide different perspectives about investigation.

In practice, however, time and budget constraints often force task teams to compromise and select the most feasible method. In the case of evaluating RBM reforms in Minas Gerais, a rigorous in-depth impact evaluation could only be approximated. The task team conducted in-depth interviews with selected government officials, analyzed quantitative aggregate data, and carried out a perception survey with public sector employees. The combined use of different methods helped yield insights about RBM reforms not achievable by one method alone. The strengths and weaknesses of each method complement the whole.

**ONLINE SURVEY**

**COST:** low/medium/high (face-to-face surveys are costly)

**STAFF HOURS:** low/medium/high

**STRENGTHS:** low cost; ease of data gathering and processing; design flexibility

**WEAKNESSES:** risk of low response rate

**TOOLS:** online survey tools (e.g., SurveyMonkey, Google Forms)

**KEY STEPS:** design and conduct questionnaire; code the answers and analyze and interpret data
REFERENCES


GOVERNANCE GLOBAL PRACTICE
Guiding Results through Public Institutions
Governance Notes captures knowledge derived from World Bank engagements and technical and financial assistance requests. The views expressed are those of the authors and do not necessarily reflect the views of the World Bank. For more information, contact: askgov@worldbank.org.