



## Improving Donor Support for Governance: The Case for More Rigorous Impact Evaluation

### Summary

There is a broad consensus in the development community on the importance of governance both as a precondition for aid and as an essential aspect for social and economic development. Each year billions of dollars are spent on supporting interventions aimed at combating corruption, promoting democracy, strengthening state institutions, upholding human rights and preserving peace.

Yet our knowledge of the effectiveness of these interventions remains very limited. Donors have made extensive use of outcome monitoring and cross-country analysis to demonstrate results in governance. Although these approaches can sketch out the progress of governance support, they do not provide convincing evidence that the observed outcome can be attributed to a given intervention alone.

What is needed is a stronger approach – an approach that measures impact rather than monitoring outcome; an approach that attributes governance outcomes to the intervention by isolating other influential factors. Such an approach is generally known as rigorous impact evaluation.

Donors have not adequately applied this approach to governance, unlike such 'traditional' sectors as health, education and labour. For instance, research shows that only one of 165 US stabilisation interventions has been rigorously evaluated. Of the 800 or so completed and ongoing rigorous impact evaluations of development programmes in the last five years, only 10 per cent or less are estimated to be governance-related.

This scenario is quite puzzling. Apart from the efficiency gains to be made from allocating scarce resources to interventions that work, conducting rigorous impact evaluation in governance is very important, given the unintended, yet potentially damaging consequences of an inappropriate governance intervention. Development cooperation in such fields as democracy, transparency and human rights, is too important, to base our knowledge purely on anecdotal evidence.

Why, then, is there so little rigorous evaluation in this field? Rigorous governance evaluation poses three key challenges:

- Governance outcomes are difficult to quantify.
- Rigorous impact evaluation can capture no more than narrow aspects of complex, 'system-wide' governance support and short-term impacts.
- A lack of incentive, since evaluation results may constrain political choices.

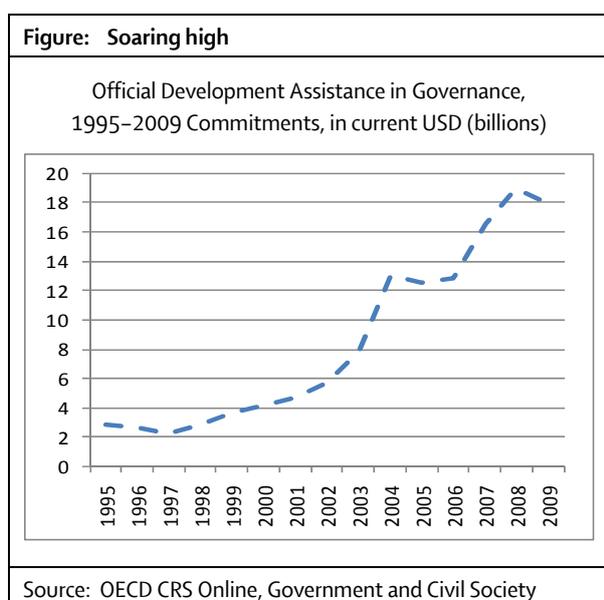
Past evaluations provide valuable recommendations to tackle these challenges. In quantifying governance outcomes, use context-specific information or conduct behavioural games. In dealing with complex 'system-wide' interventions, identify components of the programme that is viable for rigorous analysis. In improving incentives, persuade politicians to implement rigorous evaluations by studying the impact of their priority interventions and by setting appropriate expectations.

Rigorous impact evaluation in governance is difficult but feasible. As donors face increasing pressure to demonstrate results, rigorous impact evaluations should work with, not against, achieving improved governance support and development outcomes.

### Are governance interventions any good?

In the past few decades, good governance has become a core theme of international development policy. Corruption, political instability, financial mismanagement, conflict, malfunctioning legal systems, abuse of human rights and other such factors can adversely affect development and undermine any efforts to alleviate poverty. Consequently, good governance has been used both as a precondition for aid and as an essential aspect of sustainable development.

Recognising its importance, donors have actively supported governance interventions (in the form of policies, programmes and projects) in an effort to create a favourable institutional environment and well-functioning government in partner countries. Over the past decade, the world has seen a remarkable growth of governance assistance: in 2008 investment in governance reached more than USD 18.97 billion, compared to USD 2.8 billion in 1995 (see Figure).



Accompanying this large and increasing amount of governance assistance are attempts to measure its achievements. One approach is to use outcome monitoring tools, ranging from perception surveys to indices and qualitative case studies. Although monitoring tools are able to track progress and changes in outcomes, they share one striking weakness: they cannot say for certain whether and to what extent an outcome can be attributed to a given intervention. At best, they can provide descriptive findings and anecdotal evidence, but not compelling proof of impact.

Another approach is to carry out a cross-country quantitative study to analyse the impact of governance aid on development. Such studies tend, however, to suffer from simultaneous feedback from outcome to intervention and vice-versa. Governance aid may induce development, but a growing economy may also attract governance aid to support the upward trend. But which is cause and which is

effect? Even if this problem is overcome, aggregated cross-country statistics, while informative at strategic level, provide little information at the operational level of projects and programmes.

Typical evaluation reports commissioned by donors tend to ignore the attribution problem and therefore provided dubious analysis of impact. For example, a review of two dozen USAID evaluation reports on democracy and governance programmes found that 72 per cent of them failed to rule out alternative explanations that could account for the reported effects.

Hence, despite the initiatives of donors to finance and regularly monitor the progress of governance support, very little is yet known about its impact. It is imperative that donors 'do no harm' by avoiding support that could potentially weaken citizen-state relations, support based on inaccurate assumptions about the nature of local institutions and actor-constellations and support that removes pre-existing incentives for local reform processes. Relying heavily on evaluation reports using methods that do not credibly address the issue of attribution bears the risk of distorting our understanding of what works, what doesn't and why.

### The strength of rigorous impact evaluation

The lack of accuracy in pinning down impact reflects a fundamental problem in evaluation: that other factors could be responsible for the alleged outcome. Donors cannot turn back time to see what would have happened without their intervention. Nor can they easily compare the treated and non-treated individuals (or institutions) since their similarity, or lack of, must be taken into account. Rigorous impact evaluation seeks to solve these problems by finding a credible counterfactual (what would have happened without the intervention) with the use of experiments and other micro-methods (see Box 1).

#### Box 1: What is rigorous impact evaluation?

The fundamental problem in evaluation is how to establish attribution that is, to determine that the outcome is a result of the intervention and not of any other factors. In simple algebra, impact is calculated as the difference in outcome  $Y$ , with  $Y_1$  denoting the outcome if a person is exposed to the intervention and  $Y_0$  the outcome if he/she is not:

$$\text{Impact} = Y_1 - Y_0$$

At a given point in time, it is possible to observe only the outcome of the person exposed to treatment, but not the outcome when if he/she is not exposed. In other words, failure to observe both states at the same time poses a dilemma in the calculation of impact. Since it is only possible to observe one state (usually the one where the intervention took place), such quantitative approaches as experiments and quasi-experiments (an approach that mimics an experiment) can be applied to approximate the counterfactual. In general, the term "rigorous" means that the search for a valid counterfactual has been included in the analysis.

Driven by results and motivated by commitments to the Millennium Development Goals (MDGs), the Paris Declaration and the Accra Agenda for Action, many donors are now increasingly undertaking rigorous impact evaluation of development programmes. In traditional sectors, such as health, education and labour, a strong body of evidence has been produced in recent years measuring the impact of a wide range of interventions. Proponents believe that, by understanding the impact of a specific intervention – what works and why, the elusive macroeconomic effects of development assistance can eventually be identified.

While there is an overwhelming amount of rigorous impact evaluation in traditional sectors, its application in the governance arena remains limited. This finding is at odds with the widely accepted consensus that governance is crucial for development. Since the effectiveness of other development assistance and policy reforms also depends on ‘good governance’, there is an implicit need for rigour in evaluating governance support.

### Why does rigorous impact evaluation on governance continue to be a rarity?

Some argue that governance evaluation is associated with small sample size, spillover problems, a lack of good data and other such factors. While these technical issues seem to be the most obvious, rigorous governance evaluation also faces three major challenges:

- *Outcome measurement.* Governance outcomes are often difficult to measure since they are largely characterised by human behaviour, attitudes and intangible concepts. Compared with agriculture, where the effects of fertilisers are evident from farm yields, or education, where the impact of more books can be deduced from students’ test scores, measuring the impact of governance interventions on outcomes like corruption, democracy, crime and empowerment is not so straightforward. As a consequence, critics fear that the increasing focus on measuring results will lead to an agenda that can be easily quantified.
- *Inherent system-wide nature of donor support for governance.* The aim of donor support for governance is largely to improve citizen-state relations that encompass a range of actors, institutions, functions and processes. For instance, donors engaged in accountability and democratic governance provide assistance for political parties, civil society, parliaments, citizens and audit institutions and for such processes as elections, participatory budgeting and public service delivery. Given this nature, sceptics argue that rigorous impact evaluation can capture no more than narrow aspects of complex, ‘system-wide’ governance support. They also claim that this type of evaluation is incapable of depicting long-term impacts, such as institution-building and transformative processes.
- *Incentives.* Evidence may constrain political choices. Political actors and governments in partner countries may have reservations about collaborating with do-

nors to undertake rigorous evaluation of sensitive governance issues – such as poor human rights and conflict – if it can threaten their reputation, hamper their ability to make convenient decisions, discourage future funding, or instil additional conditionality.

Despite these challenges, there is considerable scope for evaluating governance interventions, perhaps far more than one would initially think. For example, recent research employing rigorous evaluation found that: audit reports helped constrain politician’s corrupt behaviour in Brazil; an anti-violence grassroots campaign reduced voter intimidation and increased voter turnout in Nigeria; and a “community-driven development” programme in post-war Sierra Leone was ineffective in transforming local institutions. Rather than focusing solely on the success or failure of the interventions, the broader message conveyed by these examples is that rigorous impact evaluation on governance is feasible. See Box 2 for a detailed example.

#### Box 2: Can grassroots participation reduce corruption?

Donor agencies have promoted numerous monitoring schemes to control corruption. Increasing grassroots participation is one of them. Citizens have the incentive to monitor their local officials since they are meant to be the direct beneficiaries of public services. But it is also subject to free-rider problem – lazy citizens share the benefits, but not the cost of monitoring – that reduces the incentive of active citizens to monitor corrupt officials. Whether or not grassroots participation is effective requires more careful investigation. With the aid of social experiments, Olken (2007) made a surprising discovery: in the context of road projects in Indonesia, audits by central government are more successful in reducing corruption than grassroots participation. In fact, grassroots participation has no impact at all. As experiments isolate other influential factors, they provide compelling evidence of what works and what doesn’t.

### Some recommendations

Given these dynamic and complex issues, conducting rigorous impact evaluation in governance can be quite demanding. While such evaluation may initially appear daunting, applying lessons from past rigorous governance evaluations and borrowing approaches from other sectors may prove to be a powerful tool for program managers:

*Although governance outcomes are difficult to measure, making use of context-specific information and/or behavioural games can help to alleviate this problem.* For example, Olken (2007) used context to construct a simple yet direct measure of corruption in the construction of village roads in Indonesia. He measured corruption by comparing the cost reported by the village and the cost estimated by an independent engineer. The difference reflected the amount of stolen money. In this case, the measure of corruption adopted is appropriate to infrastructure. Where context-specific information is used, devising an outcome measure becomes more viable than considering the general concept of corruption. In addition, some researchers conduct community games to measure behaviours that are otherwise

difficult to capture (e.g. trust and other forms of social capital). Thus, despite the outcome measurement problem, there is evidence to suggest that rigorous analysis is feasible.

*With the increasing focus on 'system-wide' governance support, identify components of the program that is viable for rigorous impact analysis.* It is important to acknowledge that there are instances where rigorous quantitative analysis is not possible; however, programme managers should not miss the opportunity to conduct one whenever it is feasible, addressing either the entire programme or certain components of it. Operationally, this suggestion implies that evaluators should be included from the inception of programme planning, and that donor staff in the governance sector should undergo training in rigorous impact evaluation. There is an urgent need for programme managers and evaluation managers to be brought together. This would balance the use of different approaches – both qualitative with quantitative – as well as improve programme design to potentially capture long-term impacts.

*Get the incentives right.* Some politicians may be persuaded to engage in rigorous impact evaluation if they are keen to know what impact their priority interventions have had. The crucial aspect is to identify these interventions and to set appropriate expectations regarding the conduct and use of the evaluation.

In general, rigorous impact evaluation may require trade-offs between costs and benefits. Clemens and Demombynes (2010) suggest that rigorous impact evaluation should be considered where the cost of collecting evidence is relatively low; where results can be made available before a policy decision; where implementing the wrong policy may be particularly harmful; where there is room to form a control group; where there is serious interest on the part of policy-makers; and where the pilot test corresponds closely to the scaled up intervention. These conditions should also apply in the governance field. If the results offer reliable knowledge, complement monitoring tools, lead to evidence-based policy-making and ultimately improve people's living conditions, the benefits of rigorous impact evaluations in governance will far outweigh its costs.

## Literature

Clemens, M. / G. Demombynes (2010): *When does rigorous impact evaluation make a difference? The case of the Millennium Villages*, Washington, DC: Center for Global Development (Working Paper 225)

Garcia, M. (2011): *Micro-methods in evaluating governance interventions*, Bonn: BMZ, (Evaluation Working Papers); online: [http://www.bmz.de/en/zentrales\\_downloadarchiv/erfolg/BMZ\\_WP\\_Micro.pdf](http://www.bmz.de/en/zentrales_downloadarchiv/erfolg/BMZ_WP_Micro.pdf)

Olken, B. (2007): *Monitoring corruption: evidence from a field experiment in Indonesia*, in: *Journal of Political Economy* 115 (2), 200–249



**Dr. Maria Melody Garcia**

Department III: Department Governance, Statehood, Security

German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE)\*)

\*) From a research project financed by the Evaluation Division of the Federal Ministry for Economic Cooperation and Development (BMZ).