

# Chapter 4

---

## Evaluation Essentials

- The majority of countries that borrowed to support reform of the core public sector had improved performance in at least some dimensions, with outliers in every category.
- IBRD countries improved more often than IDA countries, but the differences with nonborrowers were similar.
- IDA countries that had more PSR loans did better—about as well as the IBRD borrowers (one-timers)—whereas IBRD repeat borrowers did not do better.
- IEG ratings of outcome and Bank performance were also better for IBRD countries.
- Greater selectivity by countries in taking PSR loans could explain some of the difference, but IBRD would probably have better outcomes even without the selectivity effect.



Government building in Sofia, Bulgaria. Photo © J. Kaman/Travel-Images.com.

# How Public Sector Reform Outcomes Differ by Country Groups

---

**W**hen governance changes in a country, it is never possible to say precisely who is responsible—who should take credit for improvement or blame for deterioration. Of course the country itself, especially the government, has the most control and responsibility, but the degree of that control varies, as does the degree of coherence within the public sector. There is never success in PSR without favorable government involvement.

## **Measurement, Attribution, and the Role of Governments, the Bank, and Donors**

External actors come into play, including the World Bank along with other international institutions (IMF, the European Union, the United Nations Development Programme, regional development banks) and bilateral agencies. Several of them are usually involved, in close collaboration with the government when there is success, but even excellent external support alone is insufficient to guarantee success.

Measurement and timing further compound attribution. Measurements of the initial governance situation and the subsequent changes are far from perfect, even with the many improvements over the last decade. And when there is some reform effort, the effects become evident only with a lag, and an even longer lag is required to know if the effect is sustained.

Although the evidence has weaknesses, and although attribution will always be a problem, it is important to examine available evidence to con-

sider midcourse corrections if the evidence seems strong enough to suggest them.

Coordination with other donors in PSR support has gotten increasingly sophisticated and generally well adapted to the country situation. For instance, in Guyana the Inter-American Development Bank is able to do policy-based lending to complement the technical assistance work of the Bank; in Bangladesh and Indian states, the World Bank's policy-based lending complements the technical assistance grants of the United Kingdom's Department for International Development (DFID) programs. With Indian states, the Asian Development Bank and World Bank have a geographic division of labor. In Tanzania, there is now basket funding for several dimensions of PSR, to which the Bank contributes but which does not always take the lead role. The government has taken the lead, rightly insisting on better coordination. Deputy Minister Lyimo, the Bank's lead counterpart and recipient of the 2006 Gill Award, demanded of all the donors, "One process—one assessment."

Both the need for strong government participation and ownership in successful PSR and the typical participation of many donors make it difficult to have any clear attribution of results to the Bank's intervention. This is unavoidable and appropriate. What the Bank can ask for is that the process in a country to which it contributes has an impact that improves the public sector's accountability and efficiency in furthering growth and poverty reduction.

*Two available measures of outcome are changes in CPIA governance indicators and IEG project ratings.*

Two kinds of outcome measures are available for essentially all the countries receiving the Bank's lending projects for PSR: the changes in governance indicators and IEG project ratings. There is also less systematic but deeper information for a few countries in the case studies.

**Governance measures**

The CPIA is the main governance indicator considered here, though it has pros and cons (box 4.1). Indeed, one may consider the relevant CPIA indicators as the specific objectives of core PSR.<sup>1</sup> To measure changes in PSG, the analysis below uses the change of the average of CPIA indicators 13–16 (the governance CPIA) between 1999 and 2006.<sup>2</sup>

The outcomes of PSR are inherently difficult to measure. The discussion that follows provides indicative information at best. The results here indicate correlation, not causation, for several reasons:

- Imperfect measures of governance quality and the absence of these measures across wide ranges of countries and for a long enough time frame to see the effects of PSR

**Box 4.1: Pros and Cons of CPIA as a Governance Measure**

For this evaluation, two major advantages of the CPIA are that it covers essentially all borrower countries and that it rates performance categories that correspond closely to the thematic areas of the evaluation.

Bank staff make the CPIA ratings, so one must ask if this biases the ratings so as to invalidate them as a measure for this evaluation. Bank management uses the CPIA ratings to allocate budget and lending resources, especially for IDA countries, and wants them to be unbiased. An extensive benchmarking and vetting process is used to avoid bias and to counterbalance the natural tendency of country teams to make their countries look better. An external review of the CPIA in 2004 concluded that there was no serious bias (World Bank 2004a). Collier (2007) also uses the CPIA to define his category of failing states.

The team for this evaluation also tested whether more lending for a country biased its CPIA rating upward. (If there were such a bias, it would undermine the CPIA as an indicator of progress for countries that received lending for PSR.) For this test, we used the International Country Risk Guide (ICRG), one of the major external rating projects, with coverage similar to CPIA; its ratings are pos-

itively but not perfectly correlated with CPIA ratings. We ran a regression to see if a variable for past Bank lending explained the divergence between the two ratings. Although the coefficient on lending was statistically significant, it was very small—having another loan of any type increased the total CPIA by less than 0.02, on a scale of 1–6, compared with what the ICRG would predict. So for comparing groups of countries, the conclusion was that the CPIA was useable as a measure of initial conditions, and the change could measure the progress in PSR.

To interpret the results with the CPIA, one must bear in mind that the nature of progress measured by the CPIA (at least for the public sector items) evolves as one moves up the scale. The low ratings mostly refer to basic processes in areas of concern—such as having a publicly approved budget, having rules for hiring personnel and against accepting bribes, and having nominal rules for checks on executive authority. These are necessary steps to improving the public sector, but effective implementation and enforcement of rules to assure results only comes at the higher ratings. So a **3** is better than a **1** or **2**, but a citizen may not perceive any better services and accountability until the rating gets to **4** or **5**.

- The coarseness of most governance measures, in addition to uncertainty and imprecision
- Difficulty in capturing the timing of the impacts of the programs
- Nonrandom selection of countries for having PFM lending programs and governance ratings
- Omitted variable biases and the lack of information about PFM reforms with non-Bank sources of support.

### Summary Results

Three-quarters of countries getting Bank PSR lending in the period 1999–2006 experienced at least some improvement in the governance CPIA measure. In a quarter of the cases, the improvement averaged at least one notch (0.5) across all four categories, which is substantial for the relatively short period covered. Countries with PSR lending improved 0.3 points on average in the CPIA from 1999 to 2006; countries without PSR lending did not on average show a major change.

This correlation indicates a combination of two phenomena: (i) Bank support helps improve public sector performance and (ii) a selection process exists whereby countries that are more enthusiastic about PSR (and would improve somewhat anyway) are more likely to get Bank support. Both phenomena are desirable. The statistical analysis cannot tell which phenomenon predominates, and both were present in the case study countries with successful programs.

Whether countries have improved their governance also depends on income level and where they start. Countries in all categories of initial governance and IBRD/IDA were more likely to improve governance if they had a PSR project. Countries with lower initial ratings, say below 3.0, were more likely to improve than those with higher ratings. This is probably because movement in the higher ratings requires more serious changes to the way of doing business, as noted in box 4.1.

A number of countries improved even without Bank lending for PSR, especially those that started at low governance levels. Many countries had as-

sistance from other external sources, even when the Bank was not involved. And some just did it on their own.<sup>3</sup> The Bank's support for PSR is not indispensable, even though it usually does seem to be helpful. Table 4.1 indicates that when there was not lending, IDFs and AAA from the Bank were not consistently correlated with improved public sector performance.

IBRD countries with PSR projects improved more frequently than IDA countries, especially for those with mid-range initial governance ratings. In some cases, such as Cambodia and Honduras civil service, this was because project design was less well adapted to country circumstance for the IDA borrower. A higher degree of self-selection by IBRD borrowers that had PSR lending programs probably also contributed to the difference. Among countries with initially low governance (CPIA lower than 3.0), however, for reasons that are not clear, the rate of improvement for borrowers was higher for the IDA countries than the IBRD countries.

The Bank—with a variety of tools, international knowledge, and analytic capacity—has a comparative advantage for diagnosis in the technical aspects of the four PSR themes. Discussions with the government counterparts and other donors in the countries visited confirm this perception but also indicate that the extent to which this advantage is used varies across themes and across country types.

In some countries (typically IBRD countries), the government has the financial freedom and in-house technical capacity to decide whether, when, and for what it will borrow for a PSR project. Then the strategy tends to be custom made (“selective”) to the country circumstances. But in countries getting major budget support (typically IDA/PRSC), the Bank and donors more often insist on a full array of public sector reforms,

*Three-quarters of countries that received PSR lending over 1999–2006 saw an improvement in their governance CPIA measure.*

*Countries with lower initial ratings were more likely to improve than those with higher ratings.*

*IBRD countries improved more frequently than IDA countries when they had PSR loans.*

**Table 4.1: Percent of Countries with Improved CPIA Governance Scores by PSR Theme and IDA/IBRD Classification**

	IBRD		IDA or blend		Total		Major improvement
	Percent	Number	Percent	Number	Percent	Number	(>0.5) (%)
<b>Any PSR lending</b>	<b>81</b>	<b>31</b>	<b>69</b>	<b>62</b>	<b>73</b>	<b>93</b>	<b>24</b>
With $\geq 2$ PSR IL	64	11	73	30	71	41	22
With $\geq 4$ PSR AL	25	4	83	12	69	16	6
With IDF(s)	84	19	67	45	72	64	25
Without IDF	75	12	76	17	76	29	21
<b>No PSR lending</b>	<b>54</b>	<b>26</b>	<b>38</b>	<b>16</b>	<b>48</b>	<b>42</b>	<b>5</b>
With IDF(s)	44	9	50	6	47	15	7
With AAA (only)	56	16	31	13	45	29	7

Source: WB CPIA scores and IEG staff calculations.

Note: Entries show the percent and number of countries that show an improvement in the average of CPIA 13–16 between the years 1999 and 2006 (or closest year available). Columns classify countries by their 1999 IBRD/IDA classification. Rows provide this figure for subsets of countries based on the number and type of investment loans (IL) approved or active fiscal 1999–2006 and DPLs (AL) approved in fiscal 1999–2006. AAA = analytical and advisory activities; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; IDF = institutional development funds; PSR = public sector reform.

and staff often lack the time and resources to design a fully tailored product. So the result is likely to be one size fits all, off the shelf.

The relatively favorable experience with PSR in the IBRD (middle-income) countries, where the leverage of lending rarely motivates reform, shows that the Bank can motivate reforms on the basis of its high-quality expertise and advice. (See also the discussion on project ratings.) Especially in an area such as PSR, where long-term commitment is essential—one-off decisions and turnkey operations will never suffice—success has come if and only when the experts work over time with government counterparts to design and implement a project that fits local circumstances.

### **Regional differences in results**

Just as the incidence of lending varied across Regions, so did the correlation of PSR lending with changes governance scores (see table 4.2). Europe and Central Asia has the highest rate of improvement for countries getting PSR lending—90 percent—but the rate of improvement for nonborrowers is almost as high. Clearly something else is going on: European Union accession.

Almost all the countries in Europe and Central Asia not borrowing for PSR in 1999–2006 were among the first from the East to join the European Union and had done a lot of reforms with Bank support before 1999.

Latin America and the Caribbean had the second highest rate of improvement for PSR borrowers and a high differential with nonborrowers. In this Region, the improvement rate for IDA was above IBRD (both categories having significant numbers of countries). Africa and East Asia both had 70 percent improvement rates for borrowers, with Africa having the larger differential from the nonborrowers. The Middle East and North Africa and South Asia have the lowest percentages of improvement in governance CPIA scores for PSR borrowers.

### **IEG Project Ratings**

Another source of evidence on these projects is the ratings provided by IEG. Of the 238 PSR projects that closed during calendar years 1999 through 2006, three-quarters of PSR projects with IEG ratings received an overall outcome rating of at least

**Table 4.2: Percent of Countries with Improved Governance CPIA Scores by Region, 1999–2006**

Region	With Bank PSR lending		Without Bank PSR lending	
	Percent	Number	Percent	Number
Sub-Saharan Africa	70	30	47	15
East Asia and Pacific	70	10	56	9
Europe and Central Asia	90	20	86	7
Latin America and the Caribbean	75	20	25	8
Middle East and North Africa	57	7	0	2
South Asia	50	6	0	1
Total	73	93	48	42

Source: World Bank CPIA scores and IEG staff calculations.

Note: Entries show the percent and number of countries with an improvement in the average of CPIA 13–16 between 1999 and 2006 (or closest year available). Columns classify countries by their 1999 IBRD/IDA classification. Rows provide this figure for subsets of countries based on the number and type of investment loans approved or active fiscal 1999–2006 and development policy loans approved in fiscal 1999–2006.

“moderately satisfactory” and almost half received a rating of “satisfactory” or “highly satisfactory” (table 4.3).<sup>4</sup> Project design (quality-at-entry) and supervision received usually received more favorable ratings on overall outcome, which suggests that the main source of difficulty is with the countries’ performance. Another factor, however, is that project objectives (against which outcomes are judged) are sometimes overly ambitious, which pulls down the outcome ratings. More modest

objectives with the same substantive projects would have led to higher ratings.

PSR projects to IBRD countries received a larger share of “satisfactory” outcome ratings (“moderately satisfactory,” “satisfactory,” and “highly satisfactory”) than IDA and blend countries. Differences between development policy and investment lending projects were mixed but were generally small. Projects in the Europe and Central Asia Region

**Table 4.3: Summary of IEG Project Ratings for Closed PSR Projects, 1999–2006**

	Lending instrument			Lending classification			Region					
	All PSR projects	Investment	Adjustment	IBRD	Blend	IDA	Sub-Saharan Africa	East Asia and Pacific	Europe and Central Asia	Latin America and the Caribbean	Middle East and North Africa	South Asia
Overall outcome rating												
Percent S or HS	43	37	45	55	43	35	37	22	67	41	20	41
Percent MS, S, or HS	74	67	78	81	75	69	68	61	92	70	60	86
Overall Bank performance rating												
Percent S or HS	79	66	85	89	78	73	68	78	92	84	70	86
Percent MS, S, or HS	80	67	86	91	78	74	69	78	92	88	70	86

Source: IEG Project ratings database and IEG staff calculations.

Note: Table includes projects with significant PSR components that closed between January 1, 1999, and December 31, 2006, and have received IEG project ratings (238 projects). Rows indicate when cells provide the percent of projects with marginally satisfactory (MS), satisfactory (S), or highly satisfactory (HS) ratings.

performed the best on average, with South Asia and Latin America and the Caribbean Region also doing better than the others.

Projects implemented in countries with higher CPIA governance scores<sup>5</sup> received higher project ratings. For example, 67 percent of countries with a 1999 CPIA score above 4.0 received at least a “satisfactory” IEG outcome rating, whereas only 54 percent of projects implemented in countries with a 1999 CPIA score of 2.0–2.5 received this rating in the same period. In particular, projects in countries with high governance scores scored exceptionally high in the borrower preparation, implementation, and compliance ratings.

In a sense, this result is not surprising, but it raises the question of why PSR projects are not better designed and implemented in the countries that need reform most urgently—according to the Bank’s own ratings. From the case study evidence, it seems that expectations are often unrealistic and the projects sometimes fail to take care of basic matters first, especially in countries where the basics are most often missing. It is also possible that

*Projects implemented in countries with higher CPIA governance scores generally received higher project ratings.*

the government commitment is less predictable in the countries with weak governance, making it more likely that outcomes would fall below the satisfactory range even if the expected outcomes (ex ante objectives) were unbiased on average.

The success rates for PSR projects—measured by IEG ratings—was higher for IBRD loans than for IDA credits. The reasons for this are not evident in the statistics, but the country cases suggest two possible explanations. First, the design of reforms is sometimes based on models for developed countries, which are too complex for still-developing countries. And the gap is greater for the IDA countries, which tend to have institutions that are further from those of developed countries for which the models are developed.

Second, the expectations and objectives in heavy budget-support projects tend to be more ambitious and global, reflecting the donors’ list of

things that need fixing rather than the government’s list of things it is ready to do. So the reason for the lower success rate in IDA countries may have been that IBRD borrowers had a stronger say in selecting project components, including conditions.

Ratings of IDF grants, most of them for PFM, have improved overall (IEG 2007). For grants approved between fiscal 2002 and 2005, relevance was rated satisfactory in 99 percent, outcome in 79 percent, sustainability in 72 percent, Bank performance in 90 percent, and client performance in 77 percent. Relevance, sustainability, and Bank performance showed the greatest improvement since 2001. Procurement grants had the lowest performance, with less than 70 percent of the grants rated as satisfactory. Financial management grants (including auditing) performed at levels similar to the whole group; however, differences across Regions were substantial. Financial management grants included several best-practice cases, as in Thailand and Turkey.

Interviews in some country visits indicated that the government counterparts do not take IDFs as seriously as loans. Quality of supervision also remains a problem in IDF implementation. Supervision quality has depended on the availability of the task team leader and was usually better with a team leader based in the field. Continuity is also important: the few grants for which team leaders were changed more than once had lower ratings for quality of supervision (IEG 2007).

### Reasons for Country Differences

Some of the outcome differentials result from thematic factors discussed in the next chapter, but others are more cross cutting. In all four thematic areas of PSR, the Bank concentrates on the formal rules and regulations, and where divergence from actual practices is recognized, the most common strategy is to fix the formal rules in ways that encourage greater compliance and/or reduce opportunities for corruption. Such a tactic makes sense for an institution like the Bank, which works mainly with the executive branch of government. However, it is often done without much knowledge

of what created the problem in the first place and without a clear understanding of the informal processes that typically determine the outcomes.

Some country cases show that the Bank's understanding has improved regarding the difference between formal managerial processes and the practices that actually take place, driven by political economy factors. Guatemala, Bolivia, Honduras, and India are examples of relatively good practice. The Bank has to some extent taken the differences between formal process rules and administrative practices into account in designing and carrying out its support for PFM and taxes. Civil service and corruption have proven more difficult and less successful.

World Bank guidelines (World Bank 2001) recommended substantial participation by clients in PFM data gathering and analysis to facilitate ownership by clients of the results of the analysis. Since 2001, the country cases have shown a mixed record of the Bank's PERs and other PFM AAA being more responsive to demand from borrower countries, including the private sector and civil society as well as government. Bank assistance has also been more active in helping to shape the demand. Good examples were noted in Bangladesh, Tanzania, Uganda, and Vietnam. In some places, such as Bulgaria, the Russian Federation, and Ghana (until recently), the Bank had a productive interaction with the government but not with civil society.

Reasons for shortcomings include lack of incentives for Bank staff to disseminate AAA findings, Bank focus on supply-side rather than demand-side interventions, and concern among Bank staff that civil society awareness raising constitutes political lobbying, which is forbidden by the Bank's charter. This is part of a broader issue: under matrix management—according to interviews with staff—the Bank has gotten better at building cutting-edge skills, but not at integrating knowledge in support of operations at the country level.

The Bank's understanding of political economy is improving, but much of it is still at a general level,

without connection to details of the PSR agenda. Increasingly, there is agreement that a governance assessment is needed before a country proceeds to specific public sector reforms. Addressing corruption, for instance, requires understanding the nature of governance in the particular country. For this, the Bank has done much less. Through fiscal 2006 there were more than 20 IGRs, on a variety of topics, but only 5 of them gave serious attention to the political economy of the public sector as a whole. It is this sector that drives corruption and other aspects of the PSR agenda being evaluated here.

Only in a few cases (for example, Bangladesh, Bolivia, and Peru) have such analyses fed into the PSR pillar of a CAS. Some argue that this is not the comparative advantage of the Bank, given its constitutional requirement to stay out of internal politics and its dependence on the permission of governments to do its work. But the official position of the World Bank Group, reflected in presidential statements and backed by much evidence, is that fundamental improvements in PSM require political commitment and are important for growth and poverty reduction.

The Bank has done some work and could do more to understand the political foundations of governance in its partner countries. Collaborative work, especially involving local researchers, is useful in this area, and the Bank could and occasionally does take the lead in sponsoring such research. Some of this was done informally (India, Mexico, and Tanzania), but usually it is done without as much attention as to items in the regular work program.<sup>6</sup> Internal budget constraints have hindered such work in many smaller countries, where PREM staff do not have much time or money left after doing the standard macroeconomic work. As the IMF already covers that base, the Bank might consider a public sector/political economist, rather than a macroeconomist, as the core of its team in some countries.

*The Bank has improved its responsiveness to borrowers, the Bank's matrix management has caused some problems.*

*The Bank has been improving its understanding of political economy in its partner countries but needs to do more, particularly engaging local research.*