

# Are Public Financial Management Systems Improving in Low and Middle Income Countries?

## - A Preliminary Analysis based on data from PEFA assessments in 32 countries.

By FRANS E. RONSHOLT<sup>1</sup>

### ABSTRACT

*The Public Financial Management Performance Measurement Framework (better known as ‘the PEFA Framework’) was launched in June 2005 as a means to improve benchmarking and monitoring of progress of national PFM systems. Five years later this tool had been applied to national systems in 119 countries – mainly low and middle income countries. This included 32 countries where a comparison of change in systems performance over time had been undertaken, since both a baseline and at least one repeat assessment had taken place. The analysis of change in this paper utilizes a categorization of the PFM performance indicators into six groups, distinguishing formal PFM features from functional features; a methodology used in previous research on PFM systems performance in Africa, using baseline data only. Previous research analyzing patterns across a large number of countries each at one point in time found that formal features had higher scores than functional features. The present study, looking at patterns of change within countries, shows differing performance among different types of formal and functional features. It finds in particular that system features upstream and downstream in the budget management cycle are improving equally fast, whereas features that involve actor concentration and de jure reform measures are improving much faster than those involving a large number of actors across government institutions and de facto implementation of reform. These findings have wide implications for prioritization and target setting in national PFM reforms. As the database of PEFA indicator ratings used for the analysis is rapidly increasing to cover more countries and longer periods of change, the paper suggests that further research be conducted, based on a wider country sample and should consider differences in country characteristic.*

## 1. The PEFA Framework

### 1.1. Origin, purpose and content

The PEFA (Public Expenditure and Financial Accountability) Program’s <sup>2</sup> goals are to strengthen the ability of country governments and international development agencies to:

- (i) assess the condition of country public expenditure, procurement and financial accountability systems, and

---

<sup>1</sup> The author is the Head of the PEFA Secretariat, Washington DC. This paper is mainly based on the work of a team at the PEFA Secretariat, led by the author, and including Brandon Lundberg, Helena Ramos and Clay Wescott, with support from Tony Bennett, Phil Sinnott and Charles Seibert (PEFA Secretariat, 2011b).

<sup>2</sup> PEFA is a partnership program of the World Bank, the European Commission, the International Monetary Fund, the UK Department of International Development, the Swiss State Secretariat for Economic Affairs, the French Ministry of Foreign Affairs, and the Norwegian Ministry of Foreign Affairs.

(ii) develop a practical sequence of reform and capacity-building actions;  
in a manner that

- encourages country ownership,
- reduces the transaction costs to countries
- enhances donor harmonization
- allows monitoring of progress of country public financial management (PFM) performance over time
- better addresses developmental and fiduciary concerns
- leads to improved impact of reforms.

In June 2005 the Program launched the PFM Performance Measurement Framework (the PEFA Framework) in order to specifically address the first of the two goals.

### **1.2. Characteristics of the PEFA Framework**

The PEFA Framework builds on advances in theory and practice in developed and developing countries in recent decades (e.g. Caiden and Wildavsky, 1974, OECD 1995, OECD/World Bank 2003, Pollitt and Bouckaert 2004, Schiavo-Campo and Tommasi, 1999, Rosen, 2002, Rubin and Kelly 2005, Wescott, 2011). Based on this work, the outcomes of fiscal discipline, strategic resource allocation and operational efficiency in use of financial resources for service delivery can be best achieved if:

- (i) The budget is realistic and executed as intended;
- (ii) Fiscal and budget information and fiscal risk oversight is comprehensive, and open to public scrutiny;
- (iii) The budget is aligned with government policy objectives;
- (iv) Budget execution is orderly and predictable, with appropriate fiduciary controls;
- (v) Accounting and other financial reports are produced, maintained and disseminated;
- (vi) There are effective measures for external oversight.

These outcomes and six critical aspects of PFM constitute the core building blocks of the Framework. The overall structure of The PEFA Framework is presented in annex A. It has two main components:

- (i) *the PFM Performance Indicators (PI)*: a set of 31 high-level indicators – listed in annex B (and 76<sup>3</sup> sub-indicators), covering the entire budget management cycle and all main PFM subjects.
- (ii) *the PFM Performance Report (PFM-PR)*, which describes the country context; the process and scope of the assessment; the evidence for indicator ratings; and presents the integrated analytical summary.

The assessment methodology includes:

- indicator scoring on a *four-point ordinal scale*, based on: objective and transparent criteria, evidence (not perception); and is linked to the six aspects above; the scale (A, B, C, D) represents steps on the way towards application of internationally recognized, good practices;
- an *integrated analysis* to establish if the government has the tools to deliver on the three budgetary outcomes above.

---

<sup>3</sup> The first – and so far only - revision of three indicators in January 2011 increased the number of dimensions from the originally 74 to now 76.

- *focus on assessing systems performance.* It does not assess fiscal policy; nor in detail the underlying causes for good or poor performance.

### **1.3. How are the assessments implemented?**

The PEFA Framework is available as a public good. Any government or international agency may initiate and lead an assessment, and so far a range of country governments and at least twelve international agencies have done so.

Most assessments are implemented by a combination of government officials, local and international consultants as well as officials from international development agencies. The role of the PEFA Secretariat is to assist all parties involved with advice on process and technical interpretation of the Framework, as well as contributing to quality assurance in order to facilitate consistency in application of the Framework across assessments. The Secretariat also issues general guidance notes, provides training and monitors use of the Framework.

### **1.4. Global roll-out**

At March 31, 2011 a total of 227 assessment reports existed either in final or substantially complete versions<sup>4</sup>. A further 49 assessment were ongoing or planned. The completed assessments relate to 123 countries. They include assessments at the national level as well as for selected sub-national government entities – ref. application statistics in annex C. At the national government level, the coverage of baseline assessments is reaching 90% of low income countries and 75% of middle income countries, according to Lawson and Folscher, 2011.

Publication is a country level decision, and some countries prefer not to share the reports widely through open access, but share only on a need-to-know basis. More than 100 final reports are available on the Internet – all links available are found on [www.pefa.org](http://www.pefa.org)

### **1.5. Use of PEFA assessment reports**

The common information pool created by PEFA assessments have mainly been used for four purposes:

- Basis for dialogue on country PFM reform needs and priorities
- Input to operational decisions of international development agencies (e.g. provision of budget support, use/non use of government systems for aid).
- Regional peer learning
- Input to global research and evaluation.

In some regions professional networks of government PFM officials have adopted PEFA assessments as a basis for peer learning events and continuing discussions. It has been particularly valuable in regions

---

<sup>4</sup> As at October 6, 2010 the number of substantially completed and final assessments was 206, covering 119 countries.

with high coverage of PEFA assessments and where countries have a number of core characteristics in common e.g. in Eastern Europe and former Soviet states, in Francophone West and Central Africa and in the Caribbean. Often, country focused workshops discuss findings of PEFA reports, making comparison to strengths and weaknesses found for similar countries in the relevant region.

The PEFA Secretariat has organized the performance indicator ratings and their justification into a database which has been used by universities and think tanks for research purposes.

## **2. Previous research utilizing the PEFA database**

Researchers are increasingly recognizing the value of the database to explore trends in PFM systems performance in low and middle income countries – the coverage of high income countries still being quite low. Notable examples of such work include:

- Paolo de Renzio (2007) describes the main parameters that globally correlate with overall PFM systems strengths based on PEFA indicator ratings in 57 countries, and finding that PFM scores are positively associated with several factors, but in particular that variance is well explained by a combination of GDP per capita and aid dependency.
- Matthew Andrews (2009) uses the data at sub-indicator level for 31 countries to demonstrate isomorphism in PFM systems development in Africa by means of a model that distinguishes formal and functional features of PFM systems.
- Matthew Andrews (2010) builds on the 2009 paper and its data to identify PFM performance leagues of African countries and identify reform challenges for each of these leagues.
- Porter et al (2010) uses the data for African countries to identify distinguishing features of PFM systems in fragile country settings.

All of these studies utilize static information from the baseline assessments in the database.

Attempts at *dynamic* use of the database includes:

- Dorotinsky and de Renzio (2007) use Highly-Indebted Poor Countries (HIPC) assessments as a baseline, and more recent PEFA assessments in 15 HIPC countries, as the basis for tracking changes in PFM systems;
- Govt of Bangladesh and United Kingdom Department for International Development [DFID] (2007) use a retroactively established baseline to evaluate changes in the national PFM system in Bangladesh between 1992 and 2006 and relate them to PFM reform programs implemented during these 14 years. Apart from the long period involved and the retroactive aspect of the baseline assessment, this analysis is similar to the analysis of systems change that now appears in almost every repeat PEFA assessment.

PEFA Secretariat (2011) is the first publication that uses of both PEFA baseline and repeat assessments for comparison of PFM systems changes in a significant number of countries.

### **3. Analysis undertaken by the PEFA Secretariat during 2010/11**

#### **3.1. Methodology**

##### **3.1.1. Sample**

The sample of countries available for the analysis of trends is determined by the availability of repeat assessments which may be compared to a baseline. In total 45 repeat assessments had taken place as of the October 6, 2010, the cut-off date for the data used in the analysis. However, only 33 of those assessments were explicitly undertaken for the purpose of comparison with the baseline and evaluation of changes in systems performance – the other 12 assessments did not meet the criteria for being comparative<sup>5</sup>. The comparative assessments are listed in annex D.

In total 32 countries were covered (one country having two repeat assessments). About half of the countries are from Sub-Saharan Africa and there is a high level of representation of both Pacific and Caribbean island states. In addition the sample includes a handful of countries from Eastern Europe and Central Asia. This sample can hardly be considered representative of low- and middle-income countries globally. For a preliminary analysis of the potential of the database to evaluate trends in PFM systems performance, the sample was considered sufficiently large and diversified. However, a more stratified analysis based on differences between countries with specific characteristics did not yield significant results, possibly due to the limited number of countries falling in some of the categories. As the database of comparable assessments is expanding continuously, it may in the near future reach a level of comprehensiveness where such a more detailed analysis becomes meaningful.

##### **3.1.2. Period of change**

The 33 repeat assessment which were used for the analysis took place on average 33 months after the baseline, with a range from 13 (Malawi<sup>6</sup>) to 48 months (Tanzania). The standard recommendation is to undertake repeat assessments 3-5 years after the baseline, which was the case for the majority of the assessments. The timing of the assessments is often driven by a need for updates in relation to operational decisions for support by donor agencies – especially where a short interval is experienced, rather than by the need for monitoring of impact of PFM reform programs. This could suggest that the countries in the sample are more aid dependent than other low and middle income countries.

##### **3.1.3. Filtering of observation pairs**

A team of PFM experts from the PEFA Secretariat reviewed all the 33 pairs of comparative assessments, and determined for each pair of indicator observations (in total 33 x 74 = 2442 pairs) whether the

---

<sup>5</sup> A “comparative assessment” (CA) is a PEFA repeat assessment that mentions the previous assessment, compares the ratings of the two assessments, and makes some attempt to explain the differences and changes in PFM performance

<sup>6</sup> The sample includes two comparative repeat assessments for Malawi, so the interval of change in Malawi covered by the data is in fact 3 years.

comparison of scores given in the reports was robust. Two types of issues would lead to non-comparability, one being that the indicator was not scored at all in either the baseline or the repeat assessment<sup>7</sup>, the other being that other scoring issues were identified<sup>8</sup> by the team. Of all the observation pairs, 30% were found not comparable: 12% due a 'no score' in one or the other assessment, and 18% due to other scoring issues.

### **3.1.4. Classification of indicators into groups**

In addition to an analysis of improving and declining PFM elements across all the 74 PEFA indicator dimensions, a further analysis of trends utilizes a methodology (Andrews, 2009; Porter et al, 2010) that categorizes indicator dimensions in three pairs (however, the analysis “*only considers indicators/dimensions PI-5 to PI-28 as indicators PI-1 to PI-4 cover PFM system outcomes and performance and not the quality of PFM systems per se*”)<sup>9</sup>:

- *De jure* versus *de facto* establishment of a PFM system element
- PFM system elements with *actor concentration* versus *actor deconcentration*
- PFM system elements that are *upstream* in the budget management cycle versus *downstream*.

The first pair contrasts PEFA dimensions where a C or better score can be earned by a new law, or regulating a new practice, even if it is not implemented (**de jure**) with dimensions that require actual compliance, enforcement or significant engagement (**de facto**)<sup>10</sup>. For example, in the case of de jure dimension PI-11 (i), a C score is attained as long as an annual budget calendar exists, even though there may be substantial delays in implementation, with not enough time allowed to budget entities to complete detailed estimates. On the other hand, de facto dimension PI-12 (i) requires that two year forecasts of fiscal aggregates are actually produced on a rolling annual basis.

The second pair contrasts PEFA dimensions relating to budget preparation such as strategic budgeting (multi-year forecasting, strategic planning, investment planning, debt planning); annual budget preparation; legislative analysis of the annual budget; and the structure of formal budget documents on the one hand (**upstream**), and dimensions relating to budget execution such as resource management (including cash inflow and outflow management, procurement, payroll); internal control, internal audit and monitoring; accounting and reporting; external audit; and legislative analysis of audit reports on the

---

<sup>7</sup> An indicator/dimension may not have been scored for many reasons. The PEFA Secretariat developed a no scoring methodology based on the observation of three major conditions that emerged: “not rated (NR),” when insufficient information is available to score an indicator, “not applicable (NA),” when an indicator/dimension does not apply to the country being assessed, and, “not used (NU),” when there was no intention of scoring the indicator/dimension.

<sup>8</sup> Five “tracking issues” that prevented comparison were identified: (a) new evidence that was not available (or not used) for the previous assessment, (b) definition changes, (c) different interpretation of the Framework requirements, (d) different basis of sampling, and (e) the comparative assessment rescored the previous assessment to correct palpable errors. It should be emphasized that this represents a very conservative approach to ensuring comparability. For example, when the team found instances of CAs that recalculated earlier scores in order to identify change, the direction of change over time was not identified for lack of comparability.

<sup>9</sup> The full list for all coded indicators/dimensions is in Annex H.

<sup>10</sup> Indicators/dimensions were coded independently by three PFM specialists, with any disagreements discussed and reconciled (Andrews, 2009). For a full coding list, see Annex E.

other (**downstream**). The former deals with the earlier stage of the budget planning cycle, visible to donors and investors, and may be expected to be subject to less resistance than the latter aspects, which deal with controls and oversight of actual spending. For example, in the case of upstream indicator PI-5, a C score is attained as long as the budget is classified using GFS or comparable standards: a formal practice that can be monitored by donors and investors. On the other hand, downstream elements deal with more sensitive issues of managing and monitoring actual expenditures. For example, in the case of downstream indicator PI-7, a C score requires that unreported extra-budgetary expenditure be no more than 5-10 percent of total expenditure. This reduces the opportunity for non-transparent slush funds that in many countries are much greater than this amount, and may be strongly resisted by well-connected interests benefiting from such arrangements.

The third contrasts PEFA aspects under the control of central, regulatory bodies, like the Ministry of Finance (**concentrated**), with those where multiple agencies or subnational authorities need to be engaged (**deconcentrated**). For example, in the case of concentrated dimension PI-12 (ii) a C score is attained as long as a debt sustainability analysis has been undertaken during the last three years, at least for external debt: a technical job that can be done by a small team of technical staff. On the other hand, the deconcentrated dimension PI-12 (iii) requires costed sector strategies for several major sectors, which entails participation of several budget entities.

Three of the aspects can be characterized as “formal”: de jure, concentrated and upstream, while the other three considered “functional”.

This categorization was used as it was the best available and had been used in the widely published papers referenced above, and therefore important for comparability of static and dynamic use of the PEFA database. Some caution is merited, however, especially as concerns the distinction between de jure and de facto reform implementation. Many of the PEFA indicator dimensions contain both de jure and de facto aspects of implementation, with de facto aspects being increasingly important as the scores increase from C towards A. So the methodology of classifying a dimension as either de jure or de facto is rough and a future refinement warranted.

### 3.2. Findings

Previous research analyzing patterns across a large number of countries each at one point in time found that formal features had higher scores than functional features (Andrews, 2009; Porter et al, 2010). These findings may be explained as follows. Firstly, reform claims of governments are often far ahead of actual implementation. Budgeting is a political process, and announcing that reforms are underway is far easier than actually carrying them out (Wescott, 2009); thus one would expect de jure features would improve faster than de facto ones. Secondly, Woolcock and Pritchett (2002) distinguish two aspects of administrative processes: their transaction-intensity (the number of decisions that need to be made) and their discretionary character (the degree to which skilled judgment as opposed to routine procedures is needed). Fukuyama (2004) finds that the key organizational challenge is the delegation of discretion. Thus, we would expect that concentrated features would be easier to improve than deconcentrated ones. Furthermore, since upstream features would tend to involve fewer transactions and less delegation of discretion, we would also expect them to improve faster than downstream features. These hypothesized relationships between the three pairs of PFM clusters can now be tested using the PEFA dataset. The results promise to be important for monitoring the effectiveness of reforms, as well as setting reform priorities and realistic targets.

Table 1 presents the overall distribution of changes to indicator scores for the 32 countries with comparative assessments. More countries had a higher number of highest or improved scores (23 countries) than lowest or worsened scores (8 countries),<sup>11</sup> indicating a broad and welcome trend of PFM improvement across the countries surveyed. Of the latter, five were small island states with challenges not typical of the larger sample, including high vulnerability to external shocks such as hurricanes and volatile tourism earnings, severe capacity constraints with high emigration of skilled human resources, and political instability. The other three were small African states with many of the same challenges.

The suggested 'net' effect is an improvement in 18% of the indicator scores – excluding the 30% of observations where comparison was either not possible or not considered robust – a significant achievement considering the three year average interval between the baseline and repeat assessments.

Improvements in an indicator score may be one step (say from C to B) two steps (say from D to B) or three steps (from D to A). The average numbers of steps of improvement and steps of decline were almost identical at 1.36 steps, so this factor does not create a bias in comparing improvements and deterioration.

**Table 1 – Overall Distribution of Score Changes**

Change in score	Percentage of indicator dimension ratings
"A" scores maintained	11%
Increasing scores	20%
Maintained "B" or "C" scores	21%
Decreasing scores	8%
"D" scores maintained	10%
Incomparable scores (no scores and other scoring issues)	30%

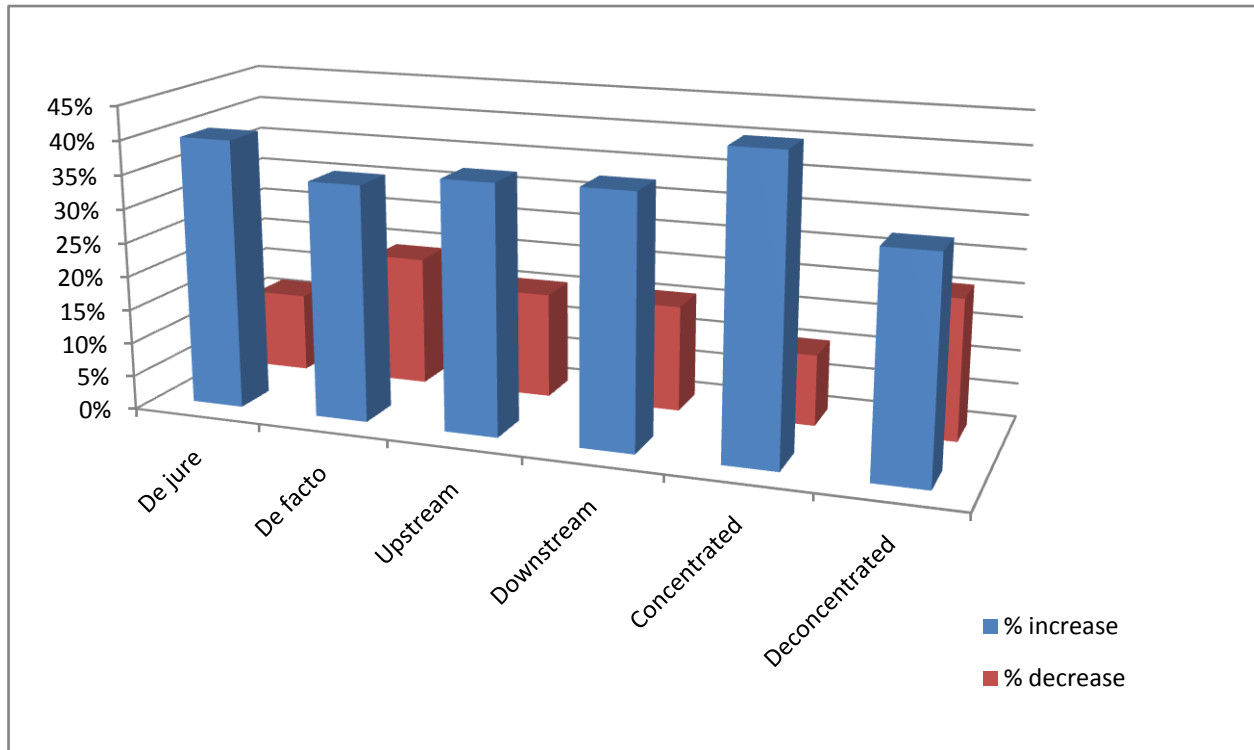
Further analysis was undertaken using the indicator categorization described in 3.1.4 above. The overall results are illustrated in diagram 1. Change is calculated as a percentage of the scores that could potentially change e.g. improvements as a percentage of all B, C and D scores (since an A score cannot increase).

---

<sup>11</sup> In two countries, the number of highest or improved scores was equal to the number of lowest or worsened scores. Highest or improved scores are combined because it's not possible to improve on an A score; lowest or worsened scores are combined because it's not possible to be worse than a D score.



**Diagram 1 - Performance changes in each of the six categories of form and function**



The general picture is one of formal features showing more progress than functional features, but also that there is markedly different progress in the three aspects of form versus function:

- *Actor concentration* showed **much higher performance** improvements than *actor deconcentration*
- *Upstream and downstream* elements of the budget cycle changed at almost **equal rate**
- *De jure* elements improved to a **moderately higher degree** than *de facto* features

Table 2 illustrates the net change (percentage improving less percentage declining) for each category.

**Table 2 – ‘Net’ performance improvement in each of the six categories**

	De jure / De facto	Upstream/ Downstream	Concentration/ De-concentrated
Form	28%	21%	33%
Function	16%	21%	11%

The three indicators/dimensions that maintained D scores in the highest percentage of cases were all examples of all three functional features (de facto, deconcentrated and downstream), where progress would be expected to be relatively difficult. In the case of the 2 indicator dimensions indicating the quality of PFM systems that worsened in the highest percentage of cases, one (PI-23) was also an example of the three functional features, while the other (PI-17 [ii]) was a case where the challenges of working downstream in the budget cycle overwhelmed the advantages of being relatively concentrated and de jure. The two dimensions that maintained A scores in the highest percentage of cases, and four that improved in the highest percentage of cases, were all examples under the centralized control of implementing bodies, where one would expect early stage success, an advantage sufficient to offset the more challenging de facto and downstream elements of some of these features.

Similar patterns can be discerned in other aspects of performance. For example, PI-25ii, *Timeliness of submission of the financial statements*, had the highest percentage of cases improving from D to A (15%), and is under the centralized control of an implementing body, as is the case with PI-13 (ii), *Taxpayer access to information on tax liabilities and administrative procedures*, that had the highest percentage of cases moving from C to A (12%), and with PI-14 (iii), *Planning and monitoring of tax audit and fraud investigation programs*, that had the highest percentage of cases moving from C to B (24%). These well performing dimensions also had some degree of correspondence with de jure and upstream features, although not as consistently. On the other hand, there was one category of good performance, the highest percentage of cases moving from D to C, where one of the two best performing dimensions was an example of three functional features (de facto, deconcentrated and downstream), PI-26 (i), *Timeliness of submission of audit reports to legislature* (18%). The other best performing dimension, PI-12 (iii), *Existence of sector strategies with multi-year costing of recurrent and investment expenditure* (both 18%), was an example of two functional features (de facto and deconcentrated). These examples show that determined governments can make headway on more challenging reforms, perhaps when there is a political opportunity or heightened support from development partners.

There is reason to be concerned about the much lower progress in improving system elements with functional features than those with formal features, because the baseline assessments show that functional features are much weaker as a starting point (ref. table 3). Especially the elements with actor deconcentration and de facto features are at risk. In contrast, downstream system elements may start from a lower base but appear to be improving as fast as the corresponding upstream elements.

**Table 3 Distribution of baseline assessment scores (comparable observations only)**

Score	De facto	Deconcentrated	Downstream	De jure	Concentrated	Upstream
A	19%	15%	18%	19%	24%	22%
B	18%	18%	22%	29%	29%	25%
C	30%	33%	33%	38%	34%	35%
D	33%	34%	28%	14%	13%	18%

Possible correlations between performance change and country characteristics were investigated in order to test the validity of previous study findings (de Renzio 2007; Andrews 2009). This analysis of correlations with a range of country characteristics resulted in no significant findings. It was concluded not to pursue multivariate regression at this stage because of the small sample size, limited range in the dependent variable, and the multitude and subjective nature of possible explanatory variables.

A possible reason for the lack of significant findings is that the country characteristics may have two contrary influences: on the one hand richer, more capable, democratic, fast growing, countries open to trade would be expected to make progress on improving PFM, but on the other hand, they may have already achieved for the same reasons a high level in the previous assessment, so further progress could be difficult. Likewise, natural resource and aid dependent countries would be expected to have less incentive for improving PFM systems, but on the other hand, they may have started from such a low level for these reasons in the previous assessment, so that they were able to show progress.

#### **4. Conclusions and Implications for country governments and international development agencies**

In conclusion, formal PFM features where progress can be achieved through adopting a new law, regulation, or technical tool, or focusing on no more than a few agencies, are more likely to improve a score than functional PFM features where progress requires actually implementing a new law or regulation, or coordinating the work of many agencies. The difference is most pronounced for PFM features where progress can be achieved working with one or a few agencies, in comparison with PFM features where many agencies are involved. Likewise, functional features are more likely to worsen a score than formal features. Both formal and functional features have higher proportions of highest and increasing scores, vs. lowest and worsening scores. These results of dynamic patterns of PEFA scores are broadly in line with static results. An important difference, however, is the finding that downstream elements are improving at the same rate as upstream features, though from a lower base, whereas the rate of progress on the other two functional features is distinctly lower than the corresponding rate of progress on the formal features. In other words, the gap between form and function is widening in those two areas. Some aspects of the results, such as the higher proportion of increasing scores than worsening scores for functional features, are encouraging, showing that even in the more difficult areas, progress is possible.

The reasons behind the trends identified here – i.e. whether trends in PFM were mainly due to the temporary capacity provided by foreign experts, or to changes in the macro environment, or to internalized government reforms – are beyond the scope of this report, but may be studied in connection with the ongoing multi-donor evaluation study of support to PFM reform. The rapidly increasing amount of data from PEFA repeat assessment would be useful for that exercise.

In setting reform priorities at country level, decision makers need to understand the effort required to implement and complete a particular type of reform and the timeframe involved. Increasingly PEFA assessments are being used in country level dialogues on PFM reform, as highlighted by Mackie and Caprio (2011). The analysis clearly shows the difference in progress between formal and functional features and suggests a need to place more emphasis on functional system features in reform plans.

International development agencies need to understand what parts of PFM systems are making progress and which ones are falling behind in order to decide where to focus their country support. The widespread perception that upstream features have been making more progress than downstream features is not supported by the findings from the present analysis. The downstream features remain weaker, but good progress is being made even if they are not catching up with upstream features. On the other hand, the poor performance in features involving actor deconcentration and de facto implementation of reforms should be a cause for concern and given attention by international development agencies due to the widening gap between systems form and function e.g. to what extent is a general focus on central finance agencies (actor concentration) an appropriate strategy, and at what stage should the emphasis shift to capacity building in the sector institutions (actor deconcentration).

## 5. Suggestions for further research

The above findings and recommendations provide an ‘appetizer’ for what the expanding database of PEFA repeat assessments can contribute to further learning about PFM performance changes. Whilst the PEFA Secretariat intends to continue some of the work, particularly as it relates to the need for development and guidance on the PEFA Framework and its application, many other opportunities are available for researchers at large. The potential for further research will also be enhanced by the growing number of comparative assessments in the database, which has increased by about 40% during the 12 months since the cut-off date for the data for the current paper.

Some suggestions for *further research* include:

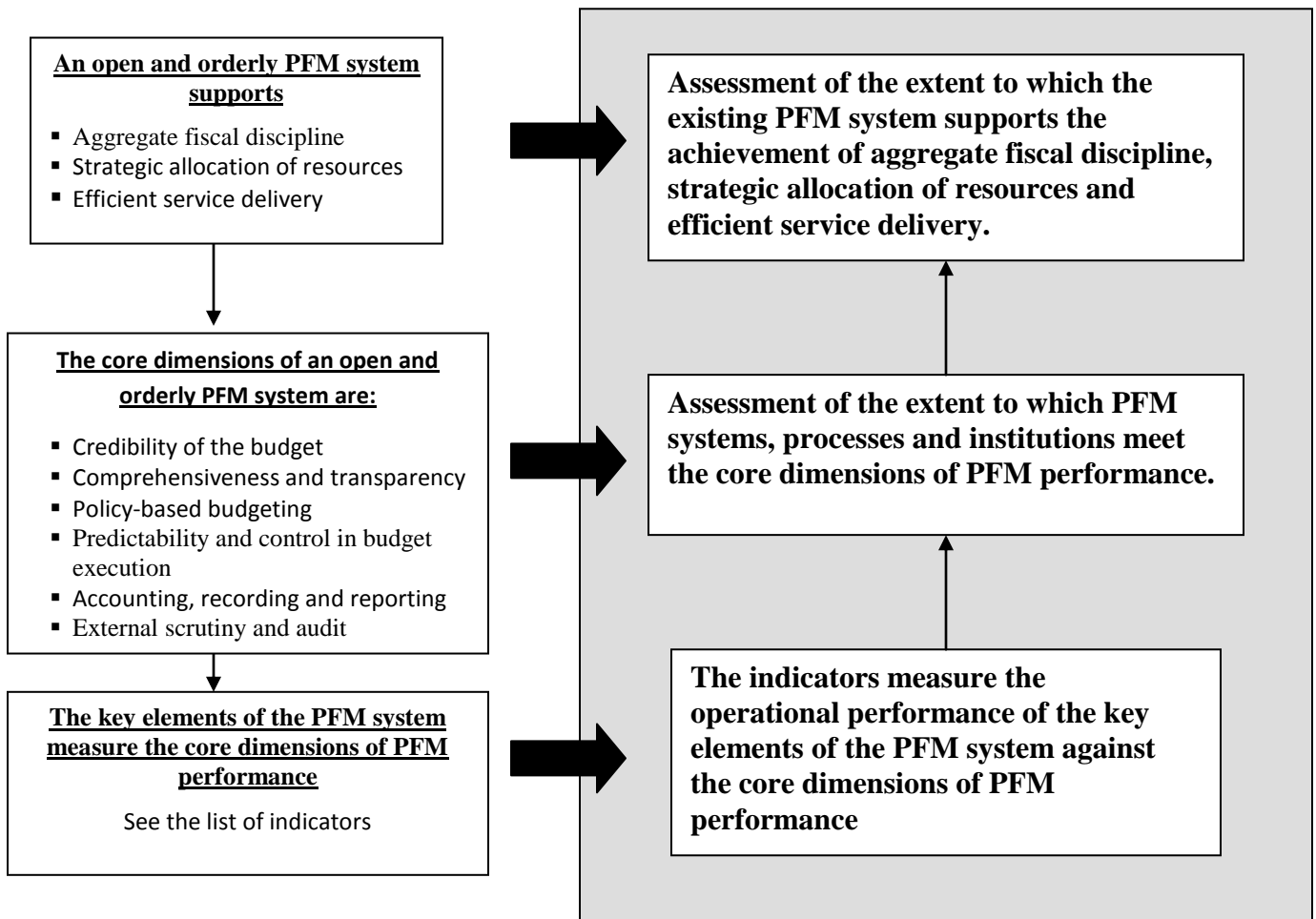
- Test to what extent findings differ in case all scored observations are included in the dataset rather than the ‘filtered’ dataset used for the present analysis; and test to what extent findings differ in case all 45 repeat assessments available as at October 2010 were used rather than the 33 comparative assessments only. This work would indicate if the Secretariat team’s two tier ‘filtering’ of available information leads to different conclusions from use of the complete raw data, and therefore suggest if similar filtering would be important for research on the expanding number of repeat assessments in the database.
- Using existing datasets, analyze factors that differentiate the 32 countries with comparative assessments from the much larger sample of countries without such. Explanatory variables could include aid (or more narrowly budget support) volumes, share of aid coming from the largest donor, share of aid coming from PEFA partner agencies, country size, democracy etc.
- On the basis of a larger number of comparative assessments now available, confirm the findings in this paper from a more diverse set of countries and with a longer average interval between the assessments and undertake analysis of correlation between country characteristics and PFM performance changes, including multivariate correlation, on this larger data set;
- Especially, investigate if any linkages between PFM performance changes and the existence of/ support to PFM reform may be identified;
- Carry out a test controlling for the possibility that more developed countries show less improvement in the comparative assessments because their scores already are high in the baseline assessment by including baseline scores as an explanatory variable in the regression;
- Refine the indicator categorization method, especially as concerns the distinction between de jure and de facto indicators, and repeat the analysis on that basis.

# ANNEXES

## A. Overall structure of the Performance Measurement Framework

### *Analytical Framework underpinning the Performance Measurement Framework*

### *The assessment provided by the Performance Measurement Framework*



## B. Overview of PEFA performance indicators <sup>12</sup>

<b>A. PFM-OUT-TURNS: Credibility of the budget</b>	
PI-1	Aggregate expenditure out-turn compared to original approved budget
PI-2	Composition of expenditure out-turn compared to original approved budget
PI-3	Aggregate revenue out-turn compared to original approved budget
PI-4	Stock and monitoring of expenditure payment arrears
<b>B. KEY CROSS-CUTTING ISSUES: Comprehensiveness and Transparency</b>	
PI-5	Classification of the budget
PI-6	Comprehensiveness of information included in budget documentation
PI-7	Extent of unreported government operations
PI-8	Transparency of inter-governmental fiscal relations
PI-9	Oversight of aggregate fiscal risk from other public sector entities.
PI-10	Public access to key fiscal information
<b>C. BUDGET CYCLE</b>	
<b>C(i) Policy-Based Budgeting</b>	
PI-11	Orderliness and participation in the annual budget process
PI-12	Multi-year perspective in fiscal planning, expenditure policy and budgeting
<b>C(ii) Predictability and Control in Budget Execution</b>	
PI-13	Transparency of taxpayer obligations and liabilities
PI-14	Effectiveness of measures for taxpayer registration and tax assessment
PI-15	Effectiveness in collection of tax payments
PI-16	Predictability in the availability of funds for commitment of expenditures
PI-17	Recording and management of cash balances, debt and guarantees
PI-18	Effectiveness of payroll controls
PI-19	Competition, value for money and controls in procurement
PI-20	Effectiveness of internal controls for non-salary expenditure
PI-21	Effectiveness of internal audit
<b>C(iii) Accounting, Recording and Reporting</b>	
PI-22	Timeliness and regularity of accounts reconciliation
PI-23	Availability of information on resources received by service delivery units
PI-24	Quality and timeliness of in-year budget reports
PI-25	Quality and timeliness of annual financial statements
<b>C(iv) External Scrutiny and Audit</b>	
PI-26	Scope, nature and follow-up of external audit
PI-27	Legislative scrutiny of the annual budget law
PI-28	Legislative scrutiny of external audit reports

<sup>12</sup> The PEFA Framework also includes 3 donor practice indicators, relevant to aid dependent countries. They are included in the analysis in the present paper which is focused only on changes in country PFM systems proper.

**C. Regional coverage table of PEFA assessment by type and region**

<i>Region</i>	<i>Baseline central government assessments</i>	<i>Baseline subnational assessments</i>	<i>Repeat assessments</i>	<i>Total Substantially Completed Reports</i>	<i>Country Coverage</i>	
					<i>Number</i>	<i>% of all countries</i>
East Asia & the Pacific	15	0	6	21	15	65%
Europe & Central Asia	16	1	4	21	16	53%
Latin American & Caribbean	26	9	10	45	27	84%
Middle East & North Africa	11	1	0	12	11	52%
Sub-Saharan Africa	41	35	26	102	42	89%
South Asia	7	7	1	15	7	88%
Other*	1	9	1	11	5	22%
Global coverage	117	62	48	227	123	67%

**D. Comparative PEFA Assessments as at October 6, 2010**

Region	Country	1st Assessment		1st Repeat Assessment		2nd Repeat Assessment	
		Lead Agency	Date of Report	Lead Agency	Date of Report	Lead Agency	Date of Report
AFR	Burkina Faso	EC	Apr. 07	Govt	Jun. 10		
AFR	Ethiopia	EC	Oct. 07	EC	Oct. 10		
AFR	Ethiopia-Benishangul Region	EC	Oct. 07	EC	Jul. 10		
AFR	Ethiopia-Harari Region	EC	Oct. 07	EC	Jul. 10		
AFR	Ethiopia-Oromiya Region	EC	Oct. 07	EC	Jul. 10		
AFR	Ghana	WB	Jun. 06	EC	Jan 10		
AFR	Guinea Bissau	WB	Jun. 06	EC	May 09		
AFR	Kenya	DFID	Jul. 06	EC	Mar. 09		
AFR	Lesotho	WB	Jun. 07	DFID	Jul. 09		
AFR	Madagascar	EC	May 06	WB	May 08		
AFR	Malawi	EC	Jul. 05	EC	Aug. 06	EC	Jun. 08
AFR	Mozambique	EC	Mar. 06	Norway	Feb. 08		
AFR	Sierra Leone	DFID	Dec. 07	DFID	Sep. 10		
AFR	Swaziland	EC	Jan. 07	WB	May 10		
AFR	Tanzania	WB	May 06	DFID	May 10		
AFR	Uganda	EC	May 06	WB	Jun. 09		
AFR	Zambia	DFID	Dec. 05	Govt	Jun. 08		
EAP	Samoa	EC	Oct. 06	Govt	Apr. 10		
EAP	Timor Leste	EC	Feb. 07	IMF	Jun. 10		
EAP	Tonga	AusAID	Sep. 07	AusAID	May 10		
EAP	Vanuatu	EC	Jul. 06	EC	Nov. 09		
ECA	Kosovo	WB	Mar. 07	Govt	May 09		
ECA	Kyrgyz Republic	DFID	Jan. 06	SECO	Dec. 09		
ECA	Moldova	EC	Apr. 06	WB	Jul. 08		
ECA	Serbia	WB	Feb. 07	Govt	Sep. 10		
LAC	Barbados	EC	Oct. 06	EC	Jul. 10		
LAC	Dominica	EC	Apr. 07	EC	Jun. 10		
LAC	Dominican Republic	EC	Nov. 07	EC	Sep. 10		
LAC	St. Kitts and Nevis	EC	Apr. 07	EC	Dec. 09		
LAC	St. Lucia	EC	Oct. 06	EC	Feb. 10		
LAC	Trinidad and Tobago	EC	Jun. 06	EC	Dec. 08		
SAR	Afghanistan	WB	Dec. 05	WB	Jun. 08		



### E. Classification of indicator dimensions according to form and function<sup>13</sup>

PEFA No.	Dimension content	De jure	De facto	Concentrated	De-concentrated	Up-stream	Down-stream
PI-5	Classification of the budget	1	0	1	0	1	0
PI-6	Comprehensiveness of information included in budget documentation	1	0	1	0	1	0
PI-7i	Level of un-reported extra-budgetary expenditure	0	1	0	1	0	1
PI-7ii	Income/expenditure information on donor-funded projects which is included in fiscal reports	0	1	0	1	0	1
PI-8i	Transparent and rules based systems for intergovernmental transfers	1	0	1	0	1	0
PI-8ii	Timeliness and reliability of budgetary information to subnational governments	0	1	0	1	1	0
PI-8iii	Collection and reporting of consolidated fiscal data for general government	0	1	0	1	0	1
PI-9i	Extent of central government monitoring of Autonomous Govt. Agencies and Public Enterprises.	0	1	0	1	0	1
PI-9ii	Extent of central government monitoring of subnational government's fiscal position	0	1	0	1	0	1
PI-10	Public access to key fiscal information	1	0	1	0	1	0
PI-11i	Existence of and adherence to a fixed budget calendar	1	0	1	0	1	0
pi-11ii	Clarity/comprehensiveness of, and political involvement in, guidance on preparing budget submissions	1	0	1	0	1	0
PI-11iii	Timely budget approval by the legislature or similarly mandated body	0	1	0	1	1	0
PI-12i	Preparation of multi -year fiscal forecasts and functional allocations	0	1	1	0	1	0
PI-12ii	Scope and frequency of debt sustainability analysis	1	0	1	0	1	0
PI-12iii	Existence of sector strategies with multi-year costing of	0	1	0	1	1	0

<sup>13</sup> Kindly provided by Matt Andrews

	recurrent and investment expenditure.						
PI-12iv	Linkages between investment budgets and forward expenditure estimates	0	1	0	1	1	0
PI-13i	Clarity and comprehensiveness of tax liabilities	1	0	1	0	0	1
PI-13ii	Taxpayer access to information on tax liabilities and administrative procedures	1	0	1	0	0	1
pi-13iii	Existence and functioning of a tax appeals mechanism	1	0	1	0	0	1
PI-14i	Controls in the taxpayer registration system.	1	0	1	0	0	1
PI-14ii	Effectiveness of penalties for non-compliance with registration and declaration obligations	0	1	1	0	0	1
PI-14iii	Planning and monitoring of tax audit and fraud investigation programs.	1	0	1	0	0	1
PI-15i	Collection of tax arrears	0	1	0	1	0	1
PI-15ii	Effectiveness of transfer of tax collections to the Treasury by the revenue administration.	0	1	1	0	0	1
PI-15iii	Frequency of Treasury accounts reconciliation between tax assessments, collections, arrears records and receipts	0	1	1	0	0	1
PI-16i	Extent to which cash flows are forecast and monitored.	1	0	1	0	0	1
PI-16ii	Information provide to Ministries, Depts. and Agencies on ceilings for expenditure commitments.	0	1	0	1	0	1
PI-16iii	Frequency and transparency of adjustments to budget allocations	0	1	0	1	0	1
PI-17i	Quality of debt data recording and reporting	1	0	1	0	0	1
PI-17ii	Extent of consolidation of the government's cash balances	1	0	1	0	0	1
PI-17iii	Systems for contracting loans and issuance of guarantees	1	0	1	0	0	1
PI-18i	Degree of integration and reconciliation between personnel records and payroll data	0	1	0	1	0	1

PI-18ii	Timeliness of changes to personnel records and the payroll	0	1	0	1	0	1
PI-18iii	Internal controls of changes to personnel records and the payroll	1	0	0	1	0	1
PI-18iv	Existence of payroll audits to identify control weaknesses and/or ghost workers	0	1	0	1	0	1
PI-19i	Evidence on the use of open competition for award of contracts	0	1	0	1	0	1
PI-19ii	Extent of justification for use of less competitive procurement methods	0	1	0	1	0	1
PI-19iii	Existence and operation of a procurement complaints mechanism	1	0	0	1	0	1
PI-20i	Effectiveness of expenditure commitment controls	0	1	0	1	0	1
PI-20ii	Comprehensiveness, relevance and understanding of other internal control rules/ procedures	1	0	0	1	0	1
PI-20iii	Degree of compliance with rules for processing and recording transactions	0	1	0	1	0	1
PI-21i	Coverage and quality of the internal audit function	0	1	0	1	0	1
PI-21ii	Frequency and distribution of reports	1	0	0	1	0	1
PI-21iii	Extent of management response to internal audit findings	0	1	0	1	0	1
PI-22i	Regularity of bank reconciliations	0	1	1	0	0	1
PI-22ii	Regularity of reconciliation and clearance of suspense accounts and advances	0	1	0	1	0	1
PI-23	Availability of information on resources received by service delivery units	0	1	0	1	0	1
PI-24i	Scope of in-year reports in terms of coverage and compatibility with budget estimates	1	0	1	0	0	1
PI-24ii	Timeliness of the issue of in-year reports	0	1	1	0	0	1
PI-24iii	Quality of information in in-year reports	0	1	1	0	0	1
PI-25i	Completeness of the financial statements	0	1	0	1	0	1

PI-25ii	Timeliness of submission of the financial statements	0	1	1	0	0	1
PI-25iii	Accounting standards used	1	0	1	0	0	1
PI-26i	Scope/nature of audit performed (incl. adherence to auditing standards)	0	1	0	1	0	1
PI-26ii	Timeliness of submission of audit reports to legislature	0	1	0	1	0	1
PI-26ii	Evidence of follow up on audit recommendations	0	1	0	1	0	1
PI-27i	Scope of the legislature's scrutiny	1	0	0	1	1	0
PI-27ii	Extent to which the legislature's procedures are well-established and respected	1	0	0	1	1	0
PI-27iii	Adequacy of time for the legislature to provide a response to budget proposals	1	0	0	1	1	0
PI-27iv	Rules for in-year amendments to the budget without ex-ante approval by the legislature	1	0	0	1	1	0
PI-28i	Timeliness of examination of audit reports by the legislature (for reports received within the last three years)	0	1	0	1	0	1
PI-28ii	Extent of hearings on key findings undertaken by the legislature	0	1	0	1	0	1
PI-28iii	Issuance of recommended actions by the legislature and implementation by the executive	0	1	0	1	0	1
<b>Totals</b>	Classification of the budget	<b>26</b>	<b>38</b>	<b>26</b>	<b>38</b>	<b>16</b>	<b>48</b>

## F. References

Allen, Schiavo-Campo & Garrity, 2004. *Assessing and Reforming Public Financial Management, A New Approach*. Washington, D.C.: The World Bank.

Andrews, Matthew 2009. *Isomorphism and the Limits to African Public Financial Management Reform*, RWP09-012. Cambridge, Mass.: Kennedy School, Harvard.

Andrews, Matthew 2010. *How Far Have Public Financial Management Reforms Come in Africa*, RWP10-018. Cambridge, Mass.: Kennedy School, Harvard.

Caiden, Naomi and Aaron B. Wildavsky, 1974. *Planning and Budgeting in poor Countries*. New York: Wiley.

de Renzio, Paulo, 2009. *Taking Stock: What do PEFA Assessments tell us about PFM systems across countries?* Working Paper 302. London: Overseas Development Institute.

de Renzio, P. and Dorotinsky, B. 2007. Tracking Progress in the Quality of PFM Systems in HIPCs: An update on past assessments using PEFA data. Washington, D.C.: World Bank.

Fukuyama, Francis 2004. "Why there is no science of public administration". *Journal of International Affairs*, 58, 1: 189-201.

Government of the Republic of Bangladesh and Department for International Development, 2007. *Assessment of the Impact of Financial Management Reforms in Bangladesh 1992 to 2006*. Dhaka: Finance Department.

Lawson, Andrew and Folscher, Alta, 2011. *Evaluation of PEFA Programme 2004 – 2010 & Development of Recommendations beyond 2011*. Oxford, UK: Fiscus Ltd and Mokoro Ltd

Mackie, Andrew and Caprio, Giovanni, 2011. *Assessing the Impact of the PEFA Framework*, A study for the PEFA Steering Committee. Washington: PEFA Secretariat.

Organization for Economic Cooperation and Development (OECD), 1995. *Budgeting for Results: Perspectives on Public Expenditure Management*. Paris: OECD.

OECD/World Bank 2003. *Budget Practices and Procedures Database*. Paris: OECD.

PEFA Secretariat, 2005. *The Public Financial Management Performance Measurement Framework*. Washington: PEFA Secretariat.

PEFA Secretariat, 2011a. *Revision of Three Performance Indicators PI-2, PI-3 and PI-19*. Washington: PEFA Secretariat.

PEFA Secretariat, 2011b. *Monitoring Report 2010 - An analysis of changes in PFM systems performance by means of PEFA repeat assessments*. Washington: PEFA Secretariat.

Pollitt, Christopher, and Geert Bouchaert. 2004. *Public Management Reform: A Comparative Analysis*. Revised edition. Oxford: Oxford University Press.

Doug Porter, Matt Andrews, Joel Turkewitz and Clay Wescott 2010. *Managing Public Finance and Procurement in Fragile and Conflicted Settings*. Washington DC: World Bank.

Rosen, Harvey. 2002. *Public Finance*. Boston, MA: McGraw-Hill.

Rubin Irene S. and Joanne Kelly. 2005. "Budget and Accounting Reforms" in Ewan Ferlie, Laurence E. Lynn, Jr. and Christopher Pollitt. *The Oxford Handbook of Public Management*. Oxford: Oxford University Press: 562-90.

Schiavo-Campo and Daniel Tommasi. 1999. *Managing Government Expenditure*. Manila: Asian Development Bank.

Wescott, Clay, 2009. "World Bank Support for Public Financial Management and Procurement: From Theory to Practice". In *Governance*. 22, 1: 139-153.

Wescott, Clay G. 2011. *Recent innovations in public financial management and procurement in Asia Pacific*. In Ari-veikko Anttiroiko and S.J. Bailey, eds. *Innovation in public governance in Asia*. Amsterdam: IOS Press.

Michael Woolcock and Lant Pritchett, 2002. *Solutions when the Solution is the Problem: Arraying the Disarray in Development*. Washington, DC: Center for Global Development.