PEFA

Public Expenditure and Financial Accountability

Dépenses publiques et responsabilité financière

PFM Performance Measurement Framework

Report on Early Experience from Application of the Framework

PEFA Secretariat

Final

November 3, 2006

TABLE OF CONTENTS

Executive Summary

Background and Purpose

i. The PEFA Program launched the PFM Performance Measurement Framework in June 2005 as part of 'the Strengthened Approach to Supporting PFM Reform'. The **objectives of the Framework** are to:

- provide reliable information on the performance of PFM systems, processes and institutions over time;
- contribute to the government reform process by determining the extent to which reforms are yielding improved performance and by increasing the ability to identify and learn from reform success;
- facilitate harmonization of the dialogue on PFM performance, reform needs and donor support between government and donors around a common PFM performance assessment and therefore contribute to reduce transaction costs for partner governments.

ii. The purpose of the present report is to capture and share the experience from the first year of applying the Framework and to identify lessons from use of the Framework and planning and managing the related processes. This should assist in planning the continued monitoring of applications and the future support of the Framework during the PEFA Program's second phase.

Extent and Characteristics of Applications

iii. **Adoption of the Framework at country level is progressing well** at a relatively steady rate of two new applications per month. Information on assessment planning suggests that by end of 2007, 70 countries may have been covered by PFM assessments based on the Framework. About 40% of the country applications completed or planned are in Sub-Saharan Africa. The rest of the assessments are fairly equally distributed across the regions of Eastern Europe & Central Asia, South Asia, East Asia & the Pacific and Latin America & the Caribbean, and to a lesser extent the Middle East & North Africa.

iv. The assessments have with one exception been led by donor agencies, with the World Bank leading in about half of the assessments, the EC in a quarter and four other agencies in the rest of the assessments. The Framework is also becoming increasingly known and accepted across donor agencies globally. **At least 19 donor agencies have actively participated in at least one of the 19 assessments reviewed.**

v. PFM assessments based on the Framework have been combined with other analytical work in two thirds of the cases and equally reported in three different formats, either as (a) stand-alone PFM-PRs, or (b) dual reporting where a PFM-PR or similar report constitutes a separate volume of a broader analytical product, or (c) integrated into a report of a broader analytical product.

vi. The reports appear to have equally aimed at three objectives being (a) informing donors' fiduciary considerations for provision and modalities of aid, (b) creating a baseline for monitoring of impact of PFM reforms, (c) contribute to formulation of PFM reforms.

Conclusions and Lessons

vii. **Most of the PFM assessments reviewed have been based on the complete set of 28 + 3 standard indicators.** Modifications or additions to the indicator set are very rare. This may reflect both adequacy of the scope of the indicator set and the substantial efforts needed to design and agree additions or changes to the standard indicators.

viii. Compliance with the Framework's principles and scoring methodology is critical to quality and transparency of the assessments and therefore, to confidence in and utilization of the assessments. With the exception of the first six assessments, **compliance has been fair with strong indications of improvements over time.** The positive trend may be attributed to a number of factors including: (a) gradual establishment of the Framework within donor organizations as the basis for PFM assessments, (b) development of support to country assessment teams from the PEFA Program, (c) upward learning curve for consultants and donor agency team leaders engaged in the assessment work. There is still substantial room for improvement, however, and this will prove crucial when comparison of ratings over time in a given country is going to take place.

ix. **Lack of adequate evidence presented in the report counted as far more important to compliance, than incorrect scoring where evidence was adequately documented.** The comprehensiveness and detail of the evidence, which is presented in the reports as justification for choosing the appropriate score, is essential for the transparency of the scoring, for the acceptance of the ratings by government and donors and therefore, for the use as a shared and agreed source of information for fiduciary and developmental purposes.

x. The Secretariat is receiving frequent requests for assistance with interpretation of individual indicators. The reports suggest a need for further guidance in order to ensure consistency in scoring. Some clarifications have already been provided on the PEFA website. **Whereas the Secretariat has already begun addressing this concern by placing some clarifications on the website, a more elaborate undertaking is required particularly with regard to those indicators which have proved to be most difficult to correctly evaluate**. There is also a need for further guidance on specific elements of the Performance Report, such as describing the public sector, the link between performance assessment and reform recommendations and the content and structure of a focused summary assessment.

xi. **Fine-tuning of the Framework is not warranted at this stage** because (a) the new and un-tested indicators introduced with the final June 2005 version of the Framework have proven to be no more difficult to assess than the indicators that were subject to testing in 2005; (b) fine-tuning would mean changing elements of the Framework and, therefore, have serious implications for consistency is scoring over time and, from the partner governments' point of view, moving the goal posts; (c) the number of assessments on which to judge a need for specific changes is too limited, and under-represents some regions with specific PFM characteristics; (d) issues of in-country comparison over time cannot yet be judged due to lack of repeat assessments.

xii. A **distinction between 'independent', 'external', 'joint government-donor', and 'government self-assessment'** was initially expected to arise in consideration of the quality and objectivity of the results. The experience shows, however, that such differences are minimal and that the best solution may be one that includes strong government involvement, multi-donor participation in planning and review as well as inputs from external or independent actors including experienced consultants.

xiii. **Partner government participation in the process is crucial for ensuring the best possible information basis for the assessment and subsequent use of the assessment for reform impact monitoring and dialogue on reform priorities**. No concerted effort has yet been made to disseminate the Framework to partner governments. Government executives on a wide scale, therefore, may not be aware of the potential benefits of the Strengthened Approach and the Framework as tools for donor harmonization and PFM reform monitoring. Government ownership of the assessment is facilitated where early orientation, training and team building is undertaken through a joint government-donor workshop, where the government designates its own counterpart team to facilitate and participate in the assessment and where assessment results and their implications are thoroughly and jointly discussed at the end of the assessment process.

xiv. **Managing expectations is important in relation to introducing a PEFA-based PFM assessment in any country.** This concerns (a) the likely level of ratings, which for most low income countries would be predominantly in the C to D range, (b) changes from earlier assessment that are likely to be modest within a 1-2 year horizon even in actively reforming countries. But expectations should also be realistic as to (c) what the PFM performance assessment can contribute to and what it cannot do.

xv. **The resources required to undertake a PEFA-based PFM assessment has often been under-estimated.** This has led to instances where inadequate information was collected to support the assessment and other cases where additional resource inputs had to be allocated mid-way through the assessment. A number of assessments have counted on a single, generalist PFM expert to undertake assessments, which has led to some PFM areas being inadequately covered. A team of two experienced assessors with complementary background, supplemented as needed by short inputs from a few specialists, appears to be a suitable formula, especially if combined with a local consultant or a government counterpart team. A total team input of 3-5 person-months should be foreseen for a central government assessment in a typical medium-sized country if implemented as a stand-alone PFM-PR.

xvi. **PEFA-based PFM assessments are often combined with broader analytical work** such as PER, CFAA, CPAR and FRA (or combinations of them) **and this can lead to tensions between different stakeholder interests** in terms of the role of the performance assessment, the time taken to finalize it, the adherence to the PEFA Framework and the links to recommendations for reform formulation and action planning. ‘Dual’ analytical products that prepare a separate PFM performance report as part of or in a coordinated parallel process with the broader analytical work offer the prospect of embracing all stakeholder interests. With appropriate packaging and sequencing of the components, such a ’dual product’ could become an important instrument in advance all of the elements of the Strengthened Approach.

xvii. **A well-defined and well-managed quality assurance mechanism is crucial for achieving a final assessment of high quality and acceptable to all stakeholders.** Such a mechanism should be agreed among all stakeholders at en early stage of planning for the assessment. The agency/person responsible for managing the mechanism should be identified and needs to systematically keep track of comments and how the report authors respond to them and reflect them in subsequent report versions. A combination of a local reference group of donor agencies and government and external reviewers from agency headquarters and/or the PEFA Secretariat appears to secure the best mix of views in terms of data reliability and compliance with the PEFA principles and methodology.

xviii. In summary, the application of the PEFA Framework for PFM performance assessment is a very decentralized process with the government and the locally involved donor agency group at the centre of planning and managing the assessment work. These stakeholders need in particular to consider and agree - ideally at an early stage in the planning process - on the following issues:

- packaging and sequencing of diagnostic products / reporting of the indicators (e.g., standalone PFM-PR, "dual products", or integration of the indicators into other products and the extent of any recommendations in any component of such a package);
- the role of various parties in conducting the assessments (ranging from government self-assessments with external validation, to assessments that are principally done by external partners with government collaboration, the leading manager of the exercise)
- the related financial and personnel resources required;
- the time for finalization of reports (taking into account the need to ensure client understanding and ownership);
- quality assurance arrangements; and
- report disclosure arrangements.

xix. There are indications that **the Framework has facilitated donor harmonization and collaboration around PFM analytical work**, but it is too early to judge to which extent the Framework has reduced the number of PFM diagnostics and promoted donor coordination as concerns reform dialogue with government and coordination of technical and financial support. There is furthermore no basis on which to judge the ability of the Framework to facilitate consistent tracking of progress in PFM performance over time.

Recommendations

xx. **All partners should focus on measures to improve quality of the assessments as the first priority.** Compliance with the principles and methodology of the PEFA Framework is key to transparency and consistency in rating and therefore to confidence in and use of the Framework.

xxi. **Donor agencies need to ensure that both head office and country level staff are well acquainted with the Framework as least for assessment planning and management purposes.** This concerns both PEFA partners and non-PEFA partners**.** The objectives of the Strengthened Approach; the role, benefits and limitations of the PFM Performance Measurement Framework; and the importance of government participation, broad donor participation, adequate assessment team resourcing and the quality review mechanism for the PFM assessment are key points to bring across to most donor agency staff.

xxii. **The benefits of the Strengthened Approach and the PFM Performance Measurement Framework should be brought across to partner governments in a much more systematic fashion that hitherto.** Government officials at executive level should be informed of the potential benefits the Approach and the Framework may offer them as tools for donor harmonization and monitoring of PFM reform impact. Officials at management level should be informed about the principles and methodology of the Framework in order to effectively participate in and contribute to the assessment work.

xxiii. **Assessors that constitute the PFM assessment teams and performance report authors need to be consistently briefed before and supported during the assessment work.** Review of completed reports have provided many clues to where difficulties arise in information collection, indicator interpretation and summarizing of assessment.

xxiv. The PEFA Program should support the above efforts by **developing a dissemination and training strategy and program** that effectively assists donor agencies, partner government and consultants along these lines.

xxv. **The PEFA Program should develop further clarification and guidance** on a continuous basis, including

(a) clarification on the role of the framework and on its limitations;

(b) clarifications to the interpretation of the indicators, especially those that have proven most difficult to assess;

(c) guidance on the evidence that would be adequate for scoring the indicators and likely sources of such evidence with emphasis on methods of collecting evidence for indicators where this has been a specific problem and on triangulating with information from non-government sources;

(d) guidance on planning and management of assessments, drawing on the examples of good practice identified so far (a first attempt in terms of checklist for assessment planning is included in Annex 5)

Such **clarification and guidance** should be made available of the website and incorporated into training programs and materials.

xxvi. **This review should be repeated in mid 2007.** A further year of experience would in particular allow more experience to be gathered on the use of the PFM performance assessments and their contributions to furthering the Strengthened Approach as well as experience from a few cases of repeat assessments. During this period the PEFA Program should also monitor utilization of the Framework at sub-national government and sector level as well as cases where non-tax revenue is an essential part of total government revenue, drawing lessons and issuing guidance in this respect.

MAIN REPORT

1. Introduction

1.1 Objective and scope of the report

1. The PEFA program is aimed at improving aid effectiveness through the Strengthened Approach to Supporting PFM Reform which has three components:

- A country-led agenda - a country led PFM reform strategy and action plan
- A coordinated program of support- an integrated, multi-year program of PFM work that supports and is aligned with the government's PFM strategy and is coordinated among the supporting donor agencies and finance institutions.
- A shared information pool – a framework for measuring results that provides consistent information on country PFM performance, including progress over time.

2. The PEFA PFM Performance Measurement Framework was developed as a tool for providing the shared pool of information as part of the Strengthened Approach. The final and official version of the PEFA PFM Performance Measurement Framework (in the following text referred to as 'the Framework') was launched in June 2005. The **objectives of the Framework** are to:

- provide reliable information on the performance of PFM systems, processes and institutions over time;
- contribute to the government reform process by determining the extent to which reforms are yielding improved performance and by increasing the ability to identify and learn from reform success;
- facilitate harmonization of the dialogue on PFM performance, reform needs and donor support between government and donors around a common PFM performance assessment and therefore contribute to reduce transaction costs for partner governments.

3. The final June 2005 version of the Framework was based on a one-year period of testing and consultation on a draft version (February 2004). During this testing period indicator scoring was applied in 24 countries and PFM-Performance Reports were prepared for more than half of those country cases. On that basis, consultations on the finalization of the Framework took place between February and June 2005.

4. During these consultations the PEFA Steering Committee decided that stock should be taken of the experience gained from application of the Framework during the 8-9 months following its official launch in order to share experiences and lessons and to assist in planning future monitoring and support of the Framework during the PEFA Program's second phase. In respect of the latter, it would be particularly important to monitor the application of the revised scoring methodology and assessment of indicators that were introduced or significantly modified as a result of the testing exercise. The present report has been prepared by the PEFA Secretariat to serve that purpose.

1.2 Sources of information for this report

5. The 19 country reports, which were received by the Secretariat during the 9½ month period from mid-June 2005 to end-March 2006 and make use of the final version of the Framework, were used as base material for this report on early experience. Seventeen of these reports are listed in Annex 1.a. with further data on each provided in Annexes 1.b, 1.c and 1.d. Two other reports are included in the background material for producing the present report, but in these cases the use of the PEFA Framework was abandoned by management decision midway in the process. For this reason, no data on those two exercises is provided in this report, but the experience from the processes that led to abandoning the PEFA element or the entire report are drawn on here and specifics are included in the aggregate information produced in the present report.

6. The 19 reports assess PFM performance at the central government level. In addition, four reports were studied in order to draw lessons on the use of the Framework at sub-national government level (ref. Annex 5). In addition to the information contained in the reviewed country reports proper, experience on process issues relating to the PFM assessments, which were ongoing at end March 2006 but had not yet produced a complete report for review, have been included in this report.

7. Further information for this report includes notes forwarded to the Secretariat by a number of assessment team leaders and lead donor representatives summarizing their experience from use of the Framework. Comments reflected both the views on the indicator set and the report content, ease/difficulties in using the Framework, and reflections on the assessment process. Such notes were received from five lead donor representatives (of which two from the World Bank) and two from consultants. The Secretariat conducted interviews with eight donor team leaders (of which five in the World Bank) and with four consultants, including the two independent consultants who have so far had the greatest exposure on the use of the Framework. Opinions were also obtained from a few donor representatives, who have followed the process of applying the Framework from a central position at headquarters. It was not deemed appropriate to conduct telephone interviews with client government representatives at this stage, and resources did not permit travel to any countries where assessments had taken place.

8. The draft report was discussed at the PEFA Steering Committee at its meeting in June 2006 and the main findings and conclusions presented to the OECD Joint Venture of PFM at its meeting later the same month. The present version of the report reflects the comments received from the participants in those events.

2. Overview of Application of the Framework

2.1 Roll-out of the Framework since its launch

9. Between June 2005 and May 2006 the Framework was adopted as a basis for or contribution to PFM analytical reports in 40 countries, of which 23 assessment exercises had reached completion of a draft or final report. Work was ongoing in the other 17 country cases. At the same time, more than 40 further PFM assessments were being planned to be implemented before the end of 2007, with use of the Framework. This means that on average 2 applications of the Framework had been substantially completed per month during the review period and suggests that the rate of roll-out may be increasing to 3 applications per months during the coming 1-2 years.

10. Of the 19 PFM assessment reports for central government assessment that were available to the Secretariat by end of March 2006, only 6 were final reports (Malawi, Zambia, Ghana, Afghanistan, Kyrgyz Republic and Mozambique), 8 were draft reports delivered to the client government (and other collaborating donors) for comment (Tanzania, Uganda, Guatemala, Panama, Bangladesh, Moldova, Fiji and Papua New Guinea), and the rest were internal draft reports not yet shared outside the leading institution(s) (e.g. Lesotho, Syria, Congo-Brazzaville).

11. The distribution of the country assessments by region and administrative heritage is shown in the following table 1. While all regions (outside the OECD countries) are represented, the Framework has clearly been most frequently used in countries with British heritage in the Sub-Saharan Africa Region. The geographical distribution of the 19 reports corresponds the distribution of the assessments expected to be undertaken during the period to the end of 2007 i.e. about 40% of the assessments in Sub-Saharan Africa, 10-20% of the assessments in each of the regions Eastern Europe & Central Asia, South Asia, East Asia & the Pacific and Latin America & the Caribbean, whereas Middle East & North Africa constitutes about 5% of the assessments.

Table 1 Distribution of PFM Assessments by region and administrative heritage[1]

Region	Central government assessments					Sub-national government assessments
	Total	British heritage	French heritage	Spanish heritage	Other heritage	
Sub-Saharan Africa	9	7	1	-	1	2
Middle East & North Africa	1	-	-	-	1	-
Europe & Central Asia	3	-	-	-	3	-
South Asia	2	1	-	-	1	1
East Asia & Pacific	2	2	-	-	-	-
Latin America & Caribbean	2	-	-	2	-	1

[1] Only including reports received by the Secretariat by end of March 2006

12. In addition to the above applications of the Framework for assessment of central government PFM performance, one assessment that substantively used the Framework has been undertaken at sub-national government level (Uganda) whilst one (Tanzania) was in progress during the preparation of this report on early experience. Information on these applications is discussed in chapter 5. The information in chapters 2, 3 and 4 concern only the assessments at central government level.

13. The timeline for implementing each of the individual country assessments is illustrated in Table 2. Several applications started with the main fact finding missions taking place before the Framework was finalized and officially launched in June 2005. This has meant that in all those cases the assessors have had to adjust the information gathered from the requirements of the 2004 draft version to the final version of the Framework after the initial data collection (e.g. Bangladesh, Afghanistan, Tanzania, Guatemala) or the assessors used the draft final version of the Framework (April 2005) which was marginally different from the final version (Fiji and Malawi).

Table 2 Timeline for PEFA Framework Applications

	2005												2006		
Country	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**
Bangladesh	<MM					D									F/CAS
Guatemala	TOR	MM			D	D									
Afghanistan		TOR	MM	D1		D2	D3			DF				F	
Tanzania	TOR		MM		D					DF					
Fiji				TOR	MM	DF									
Malawi			TOR		MM	D	F								
Ghana							TOR	MM	DF		F				
Kyrgyz							TOR		MM	D	DF		DF		F
Papua New G					TOR				MM/D						
Uganda CG							TOR				MM/D				
Uganda LG							TOR		MM	MM		D			
Zambia							TOR			MM/D		F			
Mozambique			TOR							MM	D	M/DF		M	F
Congo											TOR			MM	D
Lesotho										M			D		
Panama									TOR	MM					D
Moldova							TOR			M				M	D
Syria														MM	D

TOR	Terms of reference or concept note prepared
M or MM	Field mission or main mission
D	Draft report
DF	Draft final report
F	final report

2.2 Nature and scope of the assessment reports

14. All of the indicator-based assessments have been accompanied by narrative reports. These reports may be classified into three groups according to whether the reports (i) are ‘stand-alone’ reports that essentially follow the structure and content of a PFM Performance Report (PR), (ii) follow the structure and content of a different analytical product with the indicator-assessment incorporated, or (iii) constitutes a PFM-PR type report as a separate annex, volume or report of a broader analytical product i.e. a ‘dual’ product. An overview is given in table 3. Five of the country reports are ‘stand-alone’ PFM-PRs (even if not always given that title). Eight of the applications integrate the performance indicators into other types of reports such as combined CFAA/CPAR (Guatemala and Panama), Country Assistance Strategy and PER policy notes (Bangladesh[2]), or a PEMFAR i.e. a combination of PER/CFAA/CPAR (Afghanistan, Tanzania, Lesotho). The ‘dual’ products include Afghanistan where a PFM assessment makes up a separate volume of the five that constitutes the total analytical product. That volume is close to the scope and content of a PFM-PR. In the cases of Ghana and Moldova, fiduciary risk assessments were produced as part of the same assessment process, but presented in reports separate from the PFM-PRs. In the case of Fiji and PNG, the reports are PFM-PRs issued as separate reports on which a PER policy note and a Public Sector Reform Review drew their conclusions as regards the PFM systems.

15. With one exception, all ‘stand-alone’ PFM-PRs have been issued as consultants’ reports. The exception is Zambia, for which the PFM-PR was issued as a government document. All ‘integrated’ reports are issued as World Bank reports, and in one case as an IMF report, in collaboration with other donors and client governments as appropriate. ‘Dual’ products do not follow any specific pattern and are issued as consultants’ or donor agency reports as the organizational arrangements may have determined.

2.3 Donor agency and partner government participation

16. All of the assessments have been instigated by donor agencies. The World Bank was the leading donor in 11 of the 19 assessments; the EC led 4 assessments; while UK/DFID, Switzerland/seco and the IMF led 3 assessments in total. Sweden (not a PEFA partner) was the leading donor for one assessment (in Zambia), ref. table 2. It should be noted that the Zambia PFM assessment was led by the Zambian government, but with support from international consultants managed and financed by Sweden. As regards the assessment in Congo-Brazzaville, the EC financed and managed a consultant who produced a separate PFM-PR to form an input to an integrated analytical product led by the World Bank.

17. The six PEFA partners and eight other donor agencies have provided team members for the assessments, either from own staff resources or by funding consultants for the teams. Among non-PEFA partners, only Sweden has contributed to more than one assessment. Most assessment teams comprised either a consulting team or staff from 1-2 agencies sometimes supplemented by a consultant. There are two important exceptions. In Tanzania, the assessment team numbered 22

[2] The Framework was later integrated in the World Bank Country Assistance Strategy for Bangladesh (March 2006) whereas the policy note is yet to be completed.

and included members representing 8 donor agencies, though only 4 members were directly charged with rating the indicators. In Zambia, the government was the leading partner and provided four of the six team members, the other two being donor funded consultants.

Table 3 Leading Donors and Nature of Reports

Country	Leading donor	PFM-PR stand-alone	Dual products[3]	PFM indicators integrated into other product
Bangladesh	World Bank			X
Guatemala	World Bank			X
Afghanistan	World Bank		X	
Tanzania	World Bank			X
Fiji	World Bank		X	
Malawi	EC	X		
Ghana	DFID		X	
Kyrgyz	DFID/Switzerland	X		
Papua New Guinea	World Bank		X	
Uganda	EC	X		
Zambia	Sweden	X		
Mozambique	EC	X		
Congo	World Bank / EC		X	
Lesotho	World Bank			X
Panama	World Bank			X
Moldova	EC		X	
Syria	IMF			X

18. Apart from funding own staff inputs, funding has been provided for PFM consultants and workshops by the DFID and the EC (5 applications each), World Bank (4 applications), Sweden (3 applications), Norway, Switzerland, UNDP and the PEFA program (1 application each). Some applications were jointly funded by several donors (e.g. Uganda and Moldova), whereas others were done entirely by donor agency staff (e.g. Afghanistan). This excludes consultants who in the context of broader analytical work contributed specific sections which would not be required for a PFM assessment (e.g. free-standing procurement systems analysis).

19. Donor participation in planning, contributing information and reviewing the assessment has reached further. For example in Mozambique and Zambia, the EC and Sweden respectively

[3] 'Dual products' means a product that includes a PFM-PR or similar type of report issued as a separate volume under a common title or as a separate report under it is own title as part of a broader analytical product.

funded the consulting teams that carried out the assessments, and 16 and 10 other donors respectively contributed in the planning and review of the assessments. Globally, at least 19 donor agencies have participated in at least one of the 19 assessments.

20. Government participation has mostly been passive and reactive (ref. Annex 1.b). Only the Government of Zambia took the initiative to lead the assessment and provide most of its members, when an assessment based on the PEFA Framework was proposed by the donor group. One assessment (Ghana 2005) was carried out as a desk study and did not involve the government at all. The remaining 17 assessments were undertaken with varying degrees of government participation in planning the main mission, provision of information and commenting on the draft report. Rarely have the governments been proactive in early data collection so that time spent by consultants or donor staff in 'chasing' data could be minimized during the fact finding missions. Some governments played a more active role in the data collection and indicator rating (e.g. Fiji), even if the work was led by and the report written by a consultant.

2.4 Use of the assessments

21. The intended uses of the applications have usually fallen in three categories: (i) to inform donors on fiduciary issues related to their lending or grant aid programs, particularly in relation to budget support operations, (ii) to monitor the impact of PFM reforms, partly as a contribution to the donors' fiduciary considerations for budget support operations and partly to monitor the progress made in relation to technical assistance operations, (iii) to feed into the dialogue between government and donors on the scope, priorities and sequencing of the government's PFM reform programs or adjustments thereof. Table 4 below shows in which applications these purposes were intended according to documentation. Very few applications served only one purpose; most served two and several served all three purposes. It is likely that an additional purpose may have been served even if it was not stated; e.g. if an application serves the dialogue on PFM reform it is hard to imagine that it is not also intended to be used for monitoring performance progress.

Table 4 Purposes to be served by the PFM Assessments

Country	Donor Fiduciary Risk Consideration[4]	PFM Reform Monitoring	PFM Reform Formulation
Congo - Brazzaville	X		X
Ghana	X		
Malawi	X	X	
Mozambique	X	X	X
Lesotho			X
Tanzania	X	X	
Uganda	X	X	X
Zambia		X	X
Guatemala	X		X
Panama			X
Afghanistan			X
Bangladesh	X	X	X
Moldova	X	X	
Kyrgyz	X	X	
Fiji		X	X
PNG		X	
Syria			X

2.5 Conclusions regarding roll-out of the Framework

22. Roll-out is already progressing well at a relatively steady rate of two new applications per month. All regions are covered and assessment reports have been made available from most regions. Information on assessment planning suggests that by the end of 2007, 70 countries may have been covered by PFM assessments based on the Framework.

[4] In some cases such fiduciary considerations were included in the assessment report in others the fiduciary assessment was part of a separate report prepared as part of the same analytical work (e.g. Ghana and Moldova), whereas in yet other cases the PFM assessment informed an internal fiduciary assessment process of one or more donor agencies (e.g. Malawi).

23. The assessments have with one exception been led by donor agencies, with the World Bank leading in about half of the assessments, the EC in a quarter and four other agencies in the rest of the assessments. The Framework is also becoming increasingly known and accepted across donor agencies globally. At least 19 donor agencies have actively participated in the 19 assessments reviewed.

24. PFM assessments based on the Framework have been combined with other analytical work in two third of the cases and equally reported in three different formats, either as stand-alone PFM-PRs, or integrated into the report of a broader analytical product, or dual reporting where a PFM-PR or similar report constitutes a separate volume of a broader analytical product.

25. The reports appear to have equally focused on three objectives including (i) informing donors' fiduciary considerations for provision and chosen modalities of lending and grant aid programs, (ii) creating a baseline for monitoring of impact of PFM reforms, (iii) contribute to formulation/review of PFM reforms.

3. Utilization of and Compliance with the Framework

3.1 Use of indicator set and methodology at central government level

26. Ideally, all indicators that are applicable to the country should be covered by an assessment. The scope of work included the entire indicator set of 28+3 indicators for 14 of the 19 assessments (ref. Annex 1.d), whereas 5 assessments deliberately focused on a reduced set of indicators (e.g. Guatemala, Panama, Lesotho and Syria). None of these five assessments included donor indicators. The rationale for excluding performance indicators is not always clear but include (i) a limited scope for specific donor fiduciary considerations, (ii) indicators covered by another part of the broader analytical product (such as PI-19 on procurement) or (iii) scaling down the scope to fit the limitations of resources allocated.

27. Modification of the indicator set was explicitly attempted in two central government assessments (one substantially and the other marginally) but the modifications were eventually abandoned in both cases, and at least in one case the final assessment reverted to the full PEFA Framework. In Bangladesh, substantial modifications, based on the 2004 version of the Framework, was a result of the application already starting in 2004 when the PEFA Framework was still in the testing phase. The scoring in the draft policy note has been reworked to now present the ratings in the Country Assistance Strategy in line with the 2005 version of the Framework. The slight modifications in one application was used for background data gathering will ultimately not be included in the final report as the leading agency decided that it did not have resources to properly complete the PEFA based part of the exercise. The Afghanistan application chose to rename the 4-point ordinal scale to 4-3-2-1, but has otherwise used the Framework as intended. Several applications do not show the dimensional scores (e.g. Bangladesh, Tanzania and Lesotho) and this reduces transparency and makes verification of correct application of the Framework difficult.

28. Only the Tanzania assessment added an indicator (on corruption) for a central government assessment. However, this indicator did not follow the PEFA principles as the calibration across the 4-point scale was not disclosed. The indicator therefore lacks transparency and an agreed basis for rating.

29. Apart from the corruption indicator mentioned above (and modifications needed for assessment at district/municipality level, ref. chapter 5), assessment team leaders expressed a need for additional indicators in only one area: Administration of non-tax revenue, where such revenue constitutes a very major part of overall government revenue. Several suggestions were made to add indicators of expenditure policy (such as the protection of social budget lines etc), but this would not fit into the overall principle of the indicator set as being neutral on budget policy.

30. Not all indicators may have been scored, even if the scope of the assessment covered all of the indicators. The most common reason was lack of information on which to score the indicator. A less frequent reason was that an indicator was not considered applicable in the country being assessed. Examples of the latter reason include indicators PI-27 & 28 relating to the legislature in countries where there was no functioning legislature. D-1 (predictability of budget support) was not applied in two countries where no budget support had been provided in recent years, and one

assessment did not apply PI-8 due to very insignificant scale of financial transfers to sub-national government.

31. The Ghana report did not score 9 indicators due to lack of adequate information since it was an assessment based on secondary sources only. Similarly, the Kyrgyz assessment did not score 6 indicators. Reports did not generally explain why there was no evidence, but possible reasons are the information simply not readily available from the government, insufficient time/resources to dig out evidence, lack of knowledge in the specific subject by assessment team members, and lack of government cooperation. The table in Annex 2 shows the coverage of each indicator, i.e. the frequency by which it was scored as a percentage of the number of countries to which it was applicable. The 28 government performance indicators were scored in at least 89% of the cases, except for the revenue administration indicators (e.g. information on tax arrears often stated as not available) and PI-23 (on information on resources received by service delivery units), which were scored in 79-84% of the applicable cases. The 3 donor indicators were scored in only 63-75% of the cases.

3.2 Compliance with PEFA methodology for indicator scoring and documentation

32. Adherence to the PEFA principles and compliance with the scoring methodology[5] was assessed on the percentage of correctly applied scores that were supported by appropriate evidence in the country report. This measure focuses on compliance with use of the PEFA Framework and not on the overall quality of the assessment. Therefore, a low compliance level of a particular indicator score does not mean that the score is necessarily wrong. It means that either the data needed to score the indicator is not presented in the report, and/or the score was incorrectly decided on the basis of the data presented in the report. Compliance levels, therefore, are determined by expectations to the detail of evidence judged to be adequate by the Secretariat's reviewers; a fairly subjective standard but quite consistent for the sake of measuring relative levels of and development in compliance with the scoring methodology of the Framework.

33. In order to serve it purpose as a common information pool under the Strengthened Approach, the country report (Performance Report or equivalent) should contain all that is relevant to each country's PFM rating. Other sources of information (such as the assessors internal files) were therefore ignored for the PEFA Secretariat's judgment of compliance with the scoring methodology. One country report was not included in this analysis: Bangladesh because (i) the CAS cannot be compared to the earlier draft PFM assessment report and (ii) the CAS shows a short extract only due to the nature of this document.

34. Out of the 18 countries for which reports were reviewed by the Secretariat, 48**%** of all indicator scores were judged to be complying with PEFA scoring methodology i.e. presenting adequate evidence and correctly assigning the score on that basis. Indicators that were not scored, whether they were not applicable to the country at the time, or not scored for any other reason, were not counted in the denominator.

[5] For details on the method used in this report to measure compliance with the scoring methodology, refer to Annex 2.

35. All of the six country assessments, where the main field missions took place before the official launch of the Framework in June 2005, received compliance ratings below 50% and had an average compliance rate of 34%. Reports based on field work commencing after June 2005 had an average compliance rate of 56%. The difference in compliance rates between early and later assessments is likely to be even higher than indicated by these averages because the early assessments were all completed (or nearly) at the time of this review, whereas several of the later assessments were reviewed on the basis of their early drafts. Improvements made on those draft could increase the compliance rates significantly.

36. There was a significant difference in the compliance index for assessments presented in stand-alone PFM-PRs and PFM-PR components of a dual analytical product compared to those that integrated the Framework into the report of a different analytical product. This difference is found both for assessments started before and after June 2005 (ref. table 5). No significant differences were found between assessments led by different donor agencies within these product categories.

Table 5 Compliance index by type of analytical product

	Stand-alone PFM-PR	Dual Product	Integrated into other product
ALL: No. of assessments	5	6	7
Average compliance index	57%	57%	36%
START AFTER JUNE 2005: No. of assessments	4	4	4
Average compliance index	64%	63%	42%

37. It is important to re-emphasize that the compliance index only measures the extent to which the indicators have been scored in a transparent and methodologically correct manner. It cannot measure the overall quality of the analytical product, whether it is a PFM-PR or a PER, CFAA, PEFAR or FRA. In other words, such other analytical products can be of quality in other respects and comply with other guidelines, even if it gets a low compliance rating in the use of the PEFA Framework for scoring PFM performance indicators.

38. Nevertheless, the above differences in compliance indices point to three conclusions:

- The early applications had to refit the information collected to an indicator set that was different in the final report from when the main fact finding mission took place and, in addition, were not well supported by the PEFA program in terms of pre-mission briefings, donor staff training and additional guidance on the website.
- Where a broader analytical product is desired, the use of a dual reporting structure tends to result in a level of compliance with PEFA methodology that is almost identical to the level in stand-alone PFM-PRs.
- Analytical products that integrate the indicator set into a report format that is different from the PFM-PR, do not to the same extent focus on the indicator set and its scoring, as do the standard PFM Performance Reports, and they follow a different reporting and presentation structure. PFM-PRs are totally structured around the indicator scoring and rely on that for the analysis and conclusions, which is not the case for other products into which the Framework has been integrated.

39. The **compliance index for each indicator** showed wide variation (see table for indicators in Annex 2). Highest and lowest compliance levels were found in the indicators shown in the following table 6:

Table 6 Indicators with highest and lowest compliance to PEFA methodology

	Indicators with HIGHEST compliance	**Compliance index**
PI-6	Comprehensiveness of budget documentation	78%
PI-3	Aggregate revenue outturn compared to original approved budget	75%
PI-11	Orderliness and participation in the annual budget process	71%
	Indicators with LOWEST compliance	
PI-7	Extent of unreported government operations	6%
D-1	Predictability of direct budget support	20%
D-2	Financial information provided by donors for budgeting and reporting on project and program aid	22%

40. It is worth noting that among the entirely new indicators introduced with the 2005 version (i.e. PI-8, PI-13, PI-14, PI-15 and D-1), only indicator D-1 has given problems significantly beyond the average. The compliance indices for the other new indicators are close to the overall average (40%, 47%, 33% and 43% respectively).

41. Compliance problems in scoring an indicator may be caused by two factors: (i) adequacy of evidence presented in the report[6] and (ii) incorrect use of the evidence in determining the score. Overall, the problems of inadequate evidence (34% on average) counted as far more of an issue than incorrect scoring where evidence was sufficient (18% on average).

42. Inadequate documentation of evidence: It was expected that adequate evidence would be harder to find for some indicators than for others, particularly for the less quantitative and more judgmental indicators. On the contrary, more evidence was required for the quantitative indicators: Failure to disclose evidence, that one could reasonably believe would be available, resulted in lower compliance rating than failure to find factual support for indicators that are less subject to routine data provision. The highest degree of problems was noted with PI-7 and D-1 as detailed in table 7, both being quantitative indicators.

43. At the other end of the scale, adequate evidence was provided for 94% of the scores for PI-11 (Orderliness and participation in the budget process). Other indicators with high level evidence were PI-1, PI-3, PI-6, PI-16, PI-17, PI-19 and PI-24 for which adequate evidence was provided in

[6] In some cases, country teams explained that evidence to the required level had been collected and the indicator scores were based on this evidence, but that they did not present the evidence in the report for reasons such as client preferences/sensitivities. Without such documentation of evidence, however, the report looses its usefulness as a common information pool. These cases are therefore not distinguished from other cases of inadequate evidence.

81-83% of the cases. On average indicator scoring was based on adequate evidence in 66% of the cases.

Table 7 Indicators with a high degree of inadequate evidence

Indicator	% of scores sufficiently supported	Common problems
PI-7 Extent of unreported government operations	6%	Lack of specification of the main extra-budgetary operations, or any data on the scale of their expenditure. Lack of data on donor-funded project expenditure. Some assessors did not attempt separate ratings for the two dimensions. Where they did, dim(i)[7] was better evidenced than dim (ii)
D-1 Predictability of direct budget support	20%	Failure to obtain data on donor forecasts, or lack of forecasts, with which to calculate deviation of actual disbursements, dim(i). Greater difficulty was found with dim(ii), which involves obtaining and calculating in-year disbursement delays.

44. Scoring variance: Apart from the adequacy of evidence, assessments by the Secretariat sometimes differed from those in the country reports, using evidence presented in the report. About 20% of all scores had differences (whether upwards and downwards). The indicators that differed most frequently were PI-17 (in 44% of the cases), PI-24 and PI-25 (both in 39% of the cases). Differences arose from a number of causes, including:

- Failure to combine dimensional ratings correctly according to the appropriate methodology (M1 or M2)
- Misreading of the calibration
- Neglect of one or more of the requirements specified at each level of the calibration scale.

45. The indicators on which fewest differences in scoring were found were PI-7 (Unreported government operations, 0%), and D-1 (Predictability of direct budget support, also 0%), but there were in fact very few countries where scoring variance could be measured, as most of the assessments on these indicators were covered by lack of evidence.

46. When variance was weighted according to its magnitude (number of steps on the rating scale), the most problematic indicators were PI-1 (Aggregate expenditure outturn) and PI-2 (Composition of expenditure outturn). In a few cases, this was caused by the assessors deliberately applying a different interpretation of the indicator than the one described in the Framework, which corresponds to an implicit modification of the indicator.

47. Calibration features that contribute to scoring variance include:

[7] dim = dimension

- Real differences between requirements for scores not always being clearly identifiable (e.g. in PI-18 (iii) between a C and a D; and PI-27(iv) between B, C and D). The distinction between scores becomes ambiguous and opens the door for subjective judgment.
- Dimensions themselves being multi-dimensional (e.g. PI-17(i) and PI-24(i)). This becomes a particular problem when positive requirements for some aspects (sub-dimensions) are combined with negative requirements for others.
- Scoring a 'D' as default due to lack of data, when 'no score possible' should have been stated (e.g. Zambia and Moldova). This may happen because the lack of data is specified in a dimension as resulting in a D score for some indicators (e.g. PI-4 ii, PI-7 ii, and PI-19 i) while on the other hand the Framework (page 12) refers to the D-score being given as a residual, when no higher score can be justified. The Framework emphasizes the importance of factual evidence for each score, but could emphasize equally the importance of obtaining such evidence, and not leaving dimensions unscored, as incomplete Performance Reports are of little value.

48. The **use of 'upward arrow'** has turned out to be problematic. A few assessments use them widely or for a few selected indicators, but most assessments do not use them at all. One of the potential uses of the 'arrow' is to indicate a change in performance (since the last assessment) which did not result in a change in an indicator score. As all the assessments reviewed were baseline assessments, the 'arrow' could not be used for that purpose.

49. The use of an 'upward arrow' can be justified in cases where a PFM systems change has actually been implemented and is expected to have resulted in a performance improvement for which hard evidence is not yet available. It can be difficult to know when such a development has occurred and this makes it open for interpretation. However, the decision to use 'upward arrow' or not is an isolated issue without implications for the assignment of scores to dimensions and indicators.

50. The lack of provision for using a downward arrow makes this feature unbalanced. Some assessments have used arrows to indicate a general trend in PFM performance development, even if there has been no PEFA based assessment in the past to compare with. Comparison has then been made by links to other, earlier analytical work such as HIPC AAP assessments or CFAA or PER reports. Such cases use indicators to show upward, level or downward trend.

3.3 Adherence to Performance Report Content and Guidance

51. The **length of the PFM-PR** is specified to be up to 35 pages including only a few pages of annexes. The PFM-PRs reviewed are typically 40-50 pages including main body, summary and annexes. Four reports are significantly longer, with up to 100 pages. The PFM sections of integrated products (when constituting a separate volume such as for Afghanistan) had a substantially higher overall volume (100-400 pages). Section 3 on the indicator-led assessment often exceed the specified 18-20 pages and this mostly appears well justified in order to present detailed evidence and sources, analysis and basis of professional judgment as well as comments on reform efforts related to all 31 indicators. However, there are examples of summary sections of up

to 12 pages and the summary section in several well-written reports show that it is generally not necessary to exceed the specified 3-5 pages for a summary assessment.

52. While **statistical annexes** may add additional volume and value, some reports reproduce all or large parts of the Framework document in annexes, which of course adds significant volume. The latter has apparently been done in some cases because the Framework was not well known to the government or to the donor partners at the time. Many reports use narrative space for explaining the PEFA Framework and the guidance on each indicator. This was not the intention when proposing a recommended length, and may simply be a transitional problem as the teams are trying to explain the content of the PEFA Framework to an audience that is not yet very familiar with it. Repetition both in section 3 and in the 'summary' section often adds unnecessarily to the volume. The guidance could specify that *all* information used for rating be included in section 3, but that annexes in addition to the prescribed two may be used to elaborate further as the authors think fit, or as the study's TOR require.

53. **All sections of the PFM-PR are usually provided in the reports assessed.** Two reports, however, omitted chapter 4 on the government reform process (Malawi and Mozambique). In Malawi, it was not the intension to discuss PFM reforms in 2005 as this will be done following a 2006 repeat assessment. In Mozambique, the discussion on reform structure and management was considered better addressed by a separate analysis.

54. No **drill-down** has been provided on any specific part of the PFM scope in any PFM-PR. Broader analytical products, by their very nature, have much wider scope and often drill-down into selected PFM related areas such as procurement in Panama and human resource management of the accounting and the auditing cadre in Tanzania.

55. **Data sources** are usually provided in terms of a list of literature (government documents, previous analytical work) and a list of first-hand sources of information (interviewees) in an annex. Rarely is reference made in the narrative where the information is used for analysis. It is therefore difficult to assess on what basis a judgment or conclusion is arrived at and whose views may color the assessment. The first draft report on Mozambique included reference to the sources of information related to each indicator as part of the indicator scoring summary table. Unfortunately, this reference was removed in the final report.

56. Some indicators reflect the interface between the government and other levels of government or the non-government sector. A common weakness has been the lack of information sources among non-government institutions (chambers of commerce, taxpayers associations, tax lawyers, NGOs concerned with governance issues). The risk is that the assessment report is conveying the view of the assessed entity (whether orchestrated or by collective views) and is not sufficiently triangulating with the views of groups with different interests. In areas where assessors have not themselves been experts, this risk is particularly high.

57. Description in the **Introduction (section 1)** of structure of the public sector is rarely sufficient to understand the accountability arrangements and relationship to the central government budget. Description of sub-national governance and accountability arrangements i.e. the distinction between decentralized and deconcentrated sub-national government, is often unclear. The same is the case with the nature and importance (e.g. turnover) of autonomous government agencies (e.g.

Moldova). This has important implications for where and how SN government and AGAs are incorporated into the indicator based assessment and for the relative weight of indicators when bringing the indicator assessment results together in the summary assessment.

58. **The indicator-led assessment** (**section 3)** is the core of the report in terms of ensuring that indicator ratings are transparent and evidence based. Where necessary, this requires that any professional judgment is explained. A common reason for indicators being left unscored is uncertainty about the fit between the actual situation in the country and the calibration scenarios. This doubt arises particularly where the calibration is not (or cannot be) objective. On the other side, the available evidence may also be subjective. For instance, evidence may be based on reports of cases which are not a statistical sample and not necessarily representative of the population, such as reports by supreme audit institutions on cases of irregularity. A risk-averse assessor might conclude that there was insufficient information to rate the characteristic, and leave it unscored. A bolder assessor might make a subjective judgment, based on his/her experience. Assessors differ in their risk-aversity. The PFM-PR guidance may say: “Where there is a grey area or subjectivity in the evidence or in the calibration, the report should explicitly mention this, and offer the best estimate of the rating”. The Moldova assessment includes several examples of explicit description of the basis for a professional judgment (ref. Box I).

Box I on Good Practice
Specifying the basis for a judgment - Moldova

It is important that the reader of a PFM-PR understands the exact basis on which a particular score has been assigned. This is particularly a challenge for indicators where quantitative data is not readily available and where a judgment has to be made across a range of parallel entities or systems (e.g. across all budget entities, all payrolls, all major taxes, all functional sectors etc.). Too often, only general statements about the situation are made in the reports in a manner that fits with the indicator calibration without providing the details on how this opinion has been reached. By failing to disclose the limitations of the data, the reader can be led to believe that a score is made on firmer ground than is actually the case.

The Moldova PFM Assessment provides some examples of clearly expressed basis for assigning a score. In relation to payroll controls in indicator PI-18 dimension (ii), for example, the Moldova report specifies that “Interviews took place with the MOF and the Ministry of Education”, that “it was not possible for the practices in all Line Ministries and institutions to be examined” and that “the likelihood is that MOF & MOE may be above average performers, given their size and importance”. These considerations combined with the information obtained from the two ministries were then used for arriving at an overall rating of the dimension..

The details provided make it possible for the reader to form an opinion on the confidence in the score. It also enables a future team of assessors to collect comparable information for tracking of progress and facilitates explanation of changes in scores, if the range of data is improved at that later stage to provide a firmer basis for scoring.

59. A further problem found in many reports is that evidence is not presented comprehensively in one place but scattered between the narrative and the summary scoring table. This reduces transparency and overview of the basis for scoring.

60. Two country reports reviewed excluded section 4 on the government reform process, as already mentioned. The PFM-PR is not intended to make recommendations for reform content or management. Four reports adhered to this by not making recommendations at all (Ghana, Malawi, Congo and Kyrgyz). More often substantial volume was added by incorporating **PFM reform recommendations,** both technical recommendations for the individual PFM areas (generally incorporated into the indicator-led assessment chapter) as well as for the overall structure, priorities, sequencing and management arrangements of the reform process. While recommendations comprise a typical element in PER, CFAA and similar broader analytical work, they were also incorporated in more than half of the PFM-PRs. Incorporation of recommendations in PFM-PRs was in several cases a response to demand from the government. For example, the Zambia team was asked to propose recommendations for reform prioritization whereas the Mozambique team was asked to propose actions needed to improve the indicator scores one year on i.e. 'quick-wins'. In Uganda, it is reported that government officials were disappointed that the terms of reference for the assessment did not include specific recommendations on how to improve performance.

61. Recommendations in PFM-PRs appear in some cases well justified and formulated, whereas in other cases, the analysis is very simplistic and suggests its being added as an afterthought. In the latter case, recommendations are neither based on sufficient depth in the analysis of the underlying causes for PFM performance of the relevant part of the PFM system nor supported by analysis of the political economy of reform. These two pre-requisites for making sound reform recommendations are not called for in a standard PFM-PR, but would be, at least partially, available from broader analytical work where a PFM-PR is combined with such broader analysis.

62. While the scope of the assessment for a PFM-PR indicates that the current approach to recommendations in the PFM-PR should be maintained (i.e. not to assess the technical details of the reform program and present recommendations for its amendment), more should be done to explain the linkage between the assessment and the subsequent dialogue on the reform program content, strategy and actual effort. This linkage may need to be carefully planned at the same time as the planning of the assessment proper is taking place. In order to promote government ownership, reform recommendations could be left out of the assessment report and be discussed at an agreed workshop at a time when all parties have had an opportunity to study the assessment results and form an opinion on what reforms are required and their feasibility. Additional analytical work may then be agreed in selected priority areas for reform in order to develop a capacity building action plan.

63. **The Summary Assessment -** The guideline appear to over-emphasizes the summary against the six critical PFM dimensions, which is also the organizing framework for the indicator set, with the result that this part easily becomes a listing of indicators that score high (strengths) and low (weaknesses) without being selective on what really matters in the particular country case. Linkages in the summary to the three types of budgetary outcome are often weak. The discussion on implications for budgetary outcomes therefore often becomes a repetition of the strengths and weaknesses discussed in the previous paragraphs of the summary with little added value.

64. Analysis of budgetary outcomes is not the subject of the PFM-PR, but where such assessments are available (e.g. from PER work) it should be used to explain how the strengths/weaknesses

contribute the government ability to achieve/not-achieve budgetary outcome targets. This appears to happen only where the PFM assessment is part of a broader analytical product which explicitly includes analysis of budgetary outcomes (e.g. Afghanistan). Where PFM assessment and budget outcome analysis are included in separate analytical products, joining their findings may happen only when the last of the reports is being completed, which may not be the PFM assessment (e.g. in the case of Panama). More guidance may be needed to help assessors of PFM-PRs to create a 'story line' that emphasizes the key issues to address with options based on the information that may be available to show the linkages to budgetary outcomes.

3.4 Country specific issues

65. Two country reports (Papua New Guinea and Congo) discuss the issue of high dependence on public revenue from extractive industries. This features very prominently and explicitly in the Congo assessment, whereas in the case of PNG, it is implicit from the presentation and comment on budgetary outcomes and data for the related indicators (mainly PI-1, 2, 3 and 4) that revenue from extractive industries is important and influences budget management. The two reports, as well as discussions with some donor agency representatives, indicate that clarification is needed on how the indicator set captures non-tax revenue from extractive industries, particularly the revenue administration features that are parallel to the tax administration indicators PI-13, 14 and 15.

66. Few other specific issues are discussed in PFM Performance Reports (or equivalent). The issue of corruption and anti-corruption measures is highlighted in a couple of reports (e.g. Tanzania and Panama) and related to an attempt at scoring the issues on the same rating scale. The Bangladesh report includes a 'gap analysis' on the adoption of IPSAS standards in financial reporting, an issue of particular relevance to indicator PI-25 (on quality and timeliness of annual financial statements). The Mozambique report includes a forward estimate of indicator scores two years on, if the government implements its reform action plan as scheduled.

3.5 Consistency of assessments over time

67. **Consistency in performance assessments over time in a given country has not been possible to assess** within the limited period of this review. Several countries that had assessments in 2005 (Tanzania, Ghana and Malawi), however, are preparing a new PEFA based assessment for 2006. It is expected that initial findings in this area could be emerging towards the end of 2006, but only on a very limited scale since some of these 2006 assessments (e.g. Ghana) are considered to be baseline assessments, due to the limited scope and restricted participation of stakeholders in 2005.

68. An attempt was made in one country to **compare a previous HIPC AAP assessment with the PEFA Framework based assessment** in order to determine if there had been deterioration in PFM performance in recent years as perceived by one donor agency. This exercise proved difficult, since the data needed to assess one indicator set was only partially available in the report concerning the other indicator set. It was possible to conclude, however, that no deterioration had

been made in respect of most of the HIPC benchmarks and to identify two indicators where a drop may have occurred, but firm conclusions could not be reached on all HIPC indicators.

69. In case such a comparison is desirable in a country, the more feasible option is to make a rating according to the HIPC indicator criteria at the same time and based on the same data collection exercise as a live assessment using the new PEFA indicator set. The information that needs to be collected for the PEFA indicator set would in most cases suffice for scoring the HIPC indicators, and it will be easy to obtain any additional data for the HIPC indicators at the same time, where such data is required. Therefore, the additional staff input for such a parallel rating would be minimal. When an assessment is again undertaken using the PEFA indicator set, such a parallel rating of the HIPC indicators would no longer be needed, because tracking of progress would now be based entirely on the earlier rating of the PEFA indicator set.

70. **Cross-country comparison of performance ratings** was never an objective of the Framework, but is becoming an emerging issue. Governments that are discussing a planned PFM performance assessment are demanding information on performance in other countries they believe they are comparable to in terms of region, administrative heritage, country size and level of economic development.

71. Such cross-country comparison is not yet possible on a general scale. The reasons are that (i) most reports reviewed are still in draft stage and selected scores are likely to be changed, (ii) comparison of two countries only makes sense on indicators for which the compliance with scoring methodology is high in order to ensure consistency and adequate documentation for differences; on the current stock of reports, that limits the countries for which such a comparison can be meaningfully undertaken, (iii) comparison of a country's performance against relevant global or regional averages requires that assessments have taken place in a significant number of countries with comparable characteristics, which is not yet the case.

72. The difficulties in comparing two countries were investigated by a comparison of the assessments in Uganda and Mozambique. Although the Uganda report remained a draft version, they both had fairly high compliance rates and were undertaken under similar organizational arrangements. The choice of these two country reports was influenced by a discussion in donor agency circles of whether the relative level of scores in the two reports was justified or a result of subjectivity/inconsistency on the part of the consultant assessors. The Secretariat came to the following conclusions:

- There were 13 indicators that were correctly scored on the basis of adequate evidence in both countries. On the limited basis of these 13 indicators, there seemed to be well-justified reasons why Mozambique was scoring slightly higher on average than Uganda. Indicator PI-24 (on budget execution reports) was the only one where professional judgment could have made a difference. Even if PI-24 were more generously interpreted for Uganda, it would hardly change the overall picture of the rating levels in the two countries.
- However, the PEFA Secretariat cannot judge the correctness of the specific information/data provided in the reports for each country. The country team should in each case ensure that the information used is complete and accurate as part of their procedural requirement in any quality assurance undertaking.

- Improved presentation of evidence and related adjustment of scores for the remaining 15 indicators (8-10 in each country) could lead to entirely different conclusions.

Comparison of scoring for countries with lower scoring quality would of course be more difficult and less likely to lead to useful conclusions.

3.6 Conclusions regarding use of and compliance with the Framework

73. Compliance with the Framework's principles and scoring methodology is critical to quality and transparency of the assessments and therefore, to confidence in and utilization of the Framework. With the exception of the first six assessments, **compliance has been fair with strong indications of improvements over time.** There is still substantial room for improvement, however, and this will prove to be crucial when comparison of ratings over time in a given country is going to take place.

74. **Lack of adequate evidence presented in the report counted as far more important to compliance, than incorrect scoring where evidence was sufficient.** The comprehensiveness and detail of the evidence, which is presented in the reports as justification for choosing the appropriate score, is essential for the transparency of the scoring, for the acceptance of the ratings by government and donors and therefore, for the use as a shared and agreed source of information for fiduciary and developmental purposes.

75. **Compliance with PEFA methodology was found to be the same for stand-alone PFM Performance Reports as for PFM-PRs that were presented separately as a part of a larger analytical exercise, but significantly lower in other analytical products that integrated the performance indictors**.

76. The Secretariat is receiving frequent requests for assistance with interpretation of individual indicators. The reports suggest a need for further guidance in order to ensure consistency in scoring. Some clarifications have already been provided on the PEFA website. **Whereas the Secretariat has already began addressing this concern by placing some clarifications on the website, a more elaborate undertaking is required particularly with regard to those indicators which have proved to be most difficult to correctly evaluate**.

77. There is also **a need for further guidance on select elements of the Performance Report**, such as describing the public sector, the link between performance assessment and reform recommendations and the content and structure of a focused summary assessment.

78. **Fine-tuning of the Framework is not warranted at this stage** for the following reasons:

- The new and un-tested indicators introduced in the June 2005 version of the Framework have proven to be no more difficult to assess than the indicators that were subject to testing in 2005.
- Fine-tuning would mean changing elements of the Framework and re-issuing the Framework with serious implications for dissemination, training and, from the partner governments' point of view, moving the goal posts.

- The number of applications on which to judge a need for specific changes is too limited, and exclude or under-represent some regions with specific PFM characteristics (such as MENA, Francophone Africa and countries with high revenue dependence on extractive industries).
- Issues of in-country comparison over time cannot yet be judged due to lack of repeat assessments. This touches on the primary purpose of the Framework: to monitor progress. Some country cases for comparison over time are expected to emerge in the next year.

4. The Process of Planning and Implementing a PFM Assessment

79. The process of planning and implementing a PFM performance assessment decides participation in and organization of the work and impacts on the quality and transparency of the assessment and therefore its acceptability and utilization. This chapter summarizes the experience and lessons gathered on these issues from the year of application of the Framework (June 2005 – May 2006).

4.1 Status of the assessment

80. The status of the assessment, and specifically the government's role, has been an important issue of consideration for some PEFA partners. In principle the status options include:

- 'joint' (by government and donors collaboratively),
- 'external' (to the government),
- 'independent' (of both government and donors), or
- 'government self-assessment'

Any of these options may involve assistance by international experts and consultants.

81. Linked to this issue is the question of who owns and receives the report, and when and how to handle diverging opinions between government, donors and consultants? Do government and donors receive all drafts and briefs simultaneously, or is one party getting priority and larger influence? Are the consultants expected to submit a report based on their own 'independent' judgment with diverging government and donor views in footnotes or annex in cases where there is no consensus? Is there going to be a review by parties outside the country (quality/peer review)? If so, by whom, at what stage and what version of the report will be reviewed, and what arrangement will ensure that comments are effectively responded to in the final product?

82. With respect to the role of consultants, it is also important to consider, who controls the payment to the consultants and, therefore, the 'paymaster's' influence on the final assessment report, if it is led by a consultant.

83. The partner government, the involved donor agencies and the consultants may all have their individual incentives for deliberately or subconsciously influencing the assessment results in a specific direction. A particular solution cannot be recommended; it depends on the local stakeholders' needs, the purpose of the assessment and the trust between the parties. The experience has been that differences in assessment status have in practice been quite minor and most of the assessments have included all three groups of participants in the work. There is no suggestion that a comprehensive involvement of the government has led to lax or generous scoring, and consultants will never be more independent than determined by their client's (usually donor agency) influence on payment approval and future contract awards. Moreover, no assessment can be undertaken without agreement and participation by the government, except in isolated cases of desk assessment based on earlier analytical work (which the government will also have contributed to). With the exception of pure self-assessments (ref. Mexico states), all assessments will be joint efforts by government and donors and with consultants as leading or

supporting team members. As a good example of how such cooperation may work, see Box I on Zambia. An example of how to incorporate diverging views between the assessed government and external partners is presented in Box IV in relation to the 2006 assessment in Ghana.

Box II – Good Practice Examples
A joint assessment process led by the government – Zambia

The Strengthened Approach to Supporting PFM Reform highlights government leadership at all stages of reform preparation and implementation i.e. also of the analytical work on the basis of which reform priorities, activities and monitoring will take place.

In Zambia, a bi-annual evaluation of the Government's PFM reform program was being planned by the donor group in mid 2005. Terms of reference were produced, anticipating a team of six consultants reporting to the leading partner in the donor group (Sweden). However, the government was concerned about having another purely external evaluation of its performance in PFM (given the number of similar exercises that had been undertaken in previous years) and proposed that the evaluation should be a government-owned exercise with external facilitation. . This was accepted by the cooperating partners supporting the Public Expenditure Management and Financial Accountability (PEMFA) reform program on the basis that the evaluation should conform to the standards and procedures established by the Strengthened Approach. It was agreed that the main objective of the exercise should be to establish a baseline for monitoring progress with the PEMFA reform program in terms of performance impact; an objective directly in the government's own interest.

The government then formed its own assessment team and appointed four of its staff, mostly from the Ministry of Finance and National Planning, to undertake the assessment work. Two international consultants, recommended by the cooperating partners, of whom one had previous experience with the PEFA assessment methodology, were retained to assist and coach the government's team members. The most senior of the government officials acted as team leader.

A draft report was prepared and at that stage sent to the donor group for review. The donor group also solicited comments from the PEFA Secretariat. Some additional work still needed to be done in order to incorporate all comments received and complete the report to the satisfaction of all parties. The two international consultants were brought back to work with the government team on additional information collection and report improvements. The final report was then accepted by all parties and issued in December 2005 as a government document.

The change in approach meant that the consulting input had to be restructured to provide fewer but highly experienced consultants, more emphasis on training and coaching of the government team instead of data collection and provision of two field missions instead of one.

The Zambia government now has trained officials, who can contribute to monitoring the impact of PFM reforms, using the PEFA methodology. The government has also shown strong ownership of the assessment, and used it as a platform for dialogue concerning the reform program with the donors and other domestic stakeholders, including Parliament.

4.2 Government involvement in the assessments

84. Government participation has mostly been passive and reactive (ref. Annex 1.b). Only the Government of Zambia took the initiative to lead the assessment and provide most of its members, when an assessment based on the PEFA Framework was proposed by the donor group (ref. Box I above). One assessment (Ghana) was carried out as a desk study and did not involve the government at all.

85. The remaining 17 assessments were undertaken with varying degrees of government participation in planning the main mission, provision of information and commenting on the draft report. Rarely have the governments been proactive in early data collection so that time spent by consultants or donor staff in 'chasing' data could be minimized during the fact finding missions. Some governments played a more active role in the data collection and indicator rating (e.g. Fiji), even if the work was led by and the report written by a consultant.

86. The governments have in most cases reacted very positively to joint pre-mission planning, particularly where sensitization/training workshops on the PEFA Framework was offered early in the assessment preparation phase (e.g. Mozambique and more recently Burkina Faso) or where the PEFA Secretariat offered briefings with Q&A sessions via video-link (e.g. Kyrgyz, Zambia and Moldova). Such events have acted as important team building mechanisms across the involved government, donor agency and consultant team members and reviewers.

87. Important for the government's active participation (particularly at high level) has also been the government's incentive to ensure a fast and successful assessment e.g. where provision of budget support was directly linked to completion of the assessment (Malawi).

88. However, timing of the events has not always facilitated government participation, as they occasionally took place during periods where the government, and particularly the ministry of finance, was pre-occupied with crucial stages of the budget preparation process. E.g. workshops on the findings of the assessments in the Kyrgyz Republic and Malawi took place during the final budget preparation and parliamentary debate periods, and in Uganda the assessment field mission co-incited with the finalization of the budget guidelines.

89. The level of government involvement can be crucial for the use of the assessment. The government's leadership of the assessment in Zambia has been followed up by use of the report for dialogue on reform between government and donors and future follow-up assessments have been scheduled. The lack of government involvement in Ghana in 2005 meant that the government rejected the report as a basis for dialogue and reform monitoring, although it served for internal information purpose to a group of donors, in relation to their budget support decisions. A new PEFA based assessment was implemented in Ghana in 2006, this time with full participation of the government.

4.3 Donor coordination and participation

90. Donor participation has in many cases been coordinated through the organizational arrangement established for direct budget support or for other joint funding arrangements. As

mentioned in chapter 2, nineteen donor agencies have participated in some capacity in the first 19 assessments, six of them as leaders of the assessment or the donor group.

91. Lack of donor coordination has been an obstacle to a couple of applications where representatives from agencies who had not been associated with development of the Framework questioned its usefulness. This led to problems in getting government agreement in one case and some difficulties in obtaining information during the field mission in the other case. Eventually those problems were ironed out. These problems may have emerged because of limited dissemination of the PEFA approach to donors' field representatives and technical assistants during this early period, particularly as concerns donor agencies who are not PEFA partners.

92. Indications are that the Performance Measurement Framework has helped to improve donor coordination and collaboration around PFM assessments. The fast adoption of the Framework across countries, the many donor agencies involved and the attempts to combine the PEFA approach with previous analytical products suggest this development. A particularly good example of how the Framework brought together a large donor group is Tanzania (see Box III below). It is not possible at this stage to say to what extent the Framework has promoted donor coordination as concerns reform dialogue with government and coordination of technical and financial support as very little time to observe such effects was left for this report on early experience after completion of the PEFA based assessment in any country.

4.4 Defining the scope and modalities of the assessment

93. All World Bank analytical work is preceded by a Concept Note which is reviewed internally and with participating development partners. It is then submitted to the government for approval before the work begins in earnest. In all assessments undertaken by consultants, i.e. generally those led by the bilateral donor s and the EC, terms of reference have been prepared as a basis for contracting the consultants. Usually these terms of reference have been shared with the government and other participating donors at the country level before recruiting consultants and commencement of the work, but the extent of such consultations is not always clear. The terms of reference for the government led assessment in Zambia were also a product of a collaborative effort among the government and the 10 involved donor agencies, but had to be changed due to a change from donor to government leadership (ref. Box II).

94. Where the intention is to produce a standard PFM-PR, a reference to the strict use of the PEFA Performance Measurement Framework, for scope and methodology of the assessment and the final report, appears sufficient unless the country team wishes to deviate from the standard provisions (e.g. by including assessment of local government PFM, or progress in PFM performance since a HIPC AAP assessment). However, it may be worth emphasizing the exact extent of coverage of the assessment, since there is ambiguity in some reports as to the coverage of sub-national government. The team of assessors should be made aware that any (even minor) deviation from the Framework must be exceptional, well justified in each specific case, and approved by the supervisor(s) in advance. Implicit change of indicator calibration has occurred in a few cases, whereas explicit modifications or indicator additions have not been subject to sufficient consultation and agreement among all stakeholders.

95. It must be entirely clear to whom the team reports to and how it is to be managed; how and to whom the team issues its reports at various stages; and how it is supposed to deal with comments received from the reviewers. This should be considered along with the status of the assessment (ref. section 4.1) and quality assurance arrangements (section 4.8).

96. Annex 5 suggests a general check-list for preparation of terms of reference, and can be used equally for preparing a standard PFM-PR as for the World Bank's concept notes for broader analytical work, into which it is intended to incorporate the Framework. The European Commission has in collaboration with French Ministry of Foreign Affairs developed a model for terms of reference which is being used for PFM performance assessments led by these institutions since early 2006.

Box III – Good Practice Examples
Donor harmonization by a pro-active government – Tanzania

Tanzania has in recent years been a leading country on donor harmonization initiatives. It receives large amounts of development aid, including a large proportion in direct budget support. Some 14 donor organizations have a stake in the country's PFM system in relation to general and sector budget support operations, other use of national systems for aid provision, as well as technical assistance and investment in PFM systems development. Consequently, between 2000 and 2004, Tanzania has been subject to numerous PFM-related analytical studies, often overlapping in both timing and scope, and including PER, CFAA, CPAR, HIPC Expenditure Tracking, DFID's Fiduciary Risk Assessment, EC Compliance Test and IMF Fiscal ROSC.

During 2004 discussions within the donor PFM budget support working group took place on how to harmonize these instruments and reduce the number of assessments but no consensus was reached among them. Then in December 2004, the government sent a letter to the donors calling for future PFM assessments be subject to a 'one assessment – one process' rule i.e. the government would be willing to act as counterpart for only one PFM assessment in any one year. The donors were free to decide the scope and organization of the process, but were requested to conduct the work within current established processes.

This forced the donors to swiftly find common ground and within a couple of months the agreement was reached. The PER Annual External Review led by the World Bank in collaboration with other donors, was the most well established, comprehensive and annual assessment that had taken place in the country. It also had the most direct links to the government's leadership of the joint public expenditure work throughout the year. By adding elements of a CFAA and CPAR-updating as well as integrating the PEFA performance indicators into the PER External Review, the donor group found that all donors' information needs for their internal fiduciary assessments could be satisfied by this combined process. The combined review (the Public Expenditure and Financial Accountability Review or PEFAR) started in March 2005 and assessed central government's PFM performance on the basis of the PEFA indicators. One year later the review was repeated; this time with focus on PFM performance at local government level.

4.5 Planning the fieldwork

97. The planning phase is a period of opportunity for building a joint perception of the Framework and its use and, therefore, team spirit for the task itself, among donor agencies and the government. This is important for ensuring shared ownership of the results.

98. Effective data collection by the government itself prior to the main field mission is extremely valuable in improving the effectiveness of the main field mission, reducing the cost of an assessment and building government ownership of the process. These benefits are further enhanced if the government continues with this process and undertakes a provisional self-assessment.

99. The government's role in planning the main mission, interview schedules and data collection should be clear. It can be helpful to the assessors if the government collects most of the data of an 'objective' nature and sends it to the consultants for their review before embarking on the main mission. A preparatory visit (at least by the team leader) may also be useful in terms of identifying data gaps that need some government work before the main mission takes place. The main mission is more effective if it can concentrate on filling data gaps, interpreting and analyzing data and discussing assessment ratings between expert consultants and government/donor experts in the various subjects, rather than being spent on 'chasing' basic data. The government can support the assessors' work by designating a liaison officer or team (who have sufficient time available for the assignment).

100. The first year's experience shows that planning for 2-3 field missions for implementing the assessment is becoming recognized as highly beneficial, particularly where (i) the government is mainly a passive partner, (ii) local consultants are either not involved, or not experienced or simply not allocated enough time to follow up on data gaps etc and (iii) access to information is difficult due to size of the administration, language problems or administrative culture. These missions will typically cover (separately or combined) an initial workshop for orientation and team building among stakeholders, a main field mission followed by a preliminary report, a follow up mission for filling data gaps and discuss comments on the preliminary report, and a final workshop for presenting and discussing the results of the assessment. Allowing time and resources for these events can be important for general acceptance of the results by all stakeholders, not least the government. E.g. the Mozambique assessment involved three missions to the country despite the team having an experienced local member on site. The Mozambique government's ownership of the final product is illustrated by the publication of the report on its own website.

4.6 Resources to undertake the assessment

101. Annex 1.b shows the type and scale of resources used for the 19 reports reviewed, to the extent that such information has been made available. These resources exclude government and donor agency staff in relation to planning, supervision, review and administration of the assessment work, which some agencies may be able to quantify, but most agencies not. It has rarely been possible to separate resources used for the PFM assessment and production of a PFM-PR in the cases where this has been integrated into broader analytical work. The assessment team size has generally ranged from 1 to 9 members (when only counting those working on the PFM

assessment). Where assessment work has been undertaken mainly by a consulting team, the number of consultants has ranged from 1 to 3, with Moldova as the exception with 3 international and 2 national consultants. Tanzania is the only assessment which included a team member representing an NGO (concerned with social service provision).

102. Team members have typically comprised of economists, accountants, audit experts and procurement experts. In only one or two cases have revenue administration experts participated in the assessments, despite the inclusion of at least four revenue-focused indicators in the Framework. Procurement experts have been involved only where the assessment work has been undertaken mainly by donor agency staff (typically led by the World Bank). Procurement expertise has been very limited in teams consisting solely of a team of consultants.

103. Where data is available it suggests that 2-4 person months have typically been used to provide a full assessment of central government on the basis of the Framework (excluding donor and government staff time for planning, administration, head office management and report reviews). At the high end, 5 consulting months were used jointly by the three consultants undertaking the Mozambique assessment.

104. The size and composition of the assessment team are important parameters for ensuring a quality product. Indications are that quality is influenced by (i) number of team members (ii) number of person-months allocated (iii) experience of the team members (particularly the leader) in use of the Framework and from earlier PFM work in the country. But there are also examples that suggest that very large teams become unmanageable and tend to create internal coordination problems due to duplication of work and inconsistent approaches in the different areas covered.

105. Experience so far shows that counting on a generalist PFM expert to undertake the assessment is not the best way of ensuring quality of the overall report. The subjects covered by the PEFA indicators are so diverse that hardly any consultant will have the necessary depth of knowledge to adequately assess planning, budgeting, treasury operations, internal controls, revenue administration, payroll, procurement, accounting, auditing, parliamentary oversight, aid management, fiscal risk oversight of parastatals and local government. The team should combine knowledge that covers all of these areas. One way to achieve that would be to have two experts (including the team leader) during the entire field mission typically an economist covering planning, budget formulation and execution and an accountant covering accounting and audit, supported by a couple of experts in areas that the two main assessors do not adequately cover (e.g. procurement or revenue administration). It should be sufficient that one consultant (preferably the team leader) is experienced with the PEFA methodology, but this can be an important feature for a team since the compliance ratings show a steady upward trend (upward learning curve) for those, admittedly few, consultants who have been team leaders or dominant team members in more than one assessment. A local consultant may also be an asset to the team, particularly to assist in the initial data collection, mission preparation and any data follow-up after the main mission.

106. The resource allocation needed to implement an assessment depends on a number of factors such as

- collaboration by the government in terms of pre-mission data collection, initial self-assessment etc,

- recent PFM analytical work available
- language/translation problems
- scope of the TOR in terms of coverage of central/general government and inclusion of pre-mission and results presentation workshops
- size of country/geographical concentration of informants
- general experience and PFM subject matter coverage of the team members/consultants,
- the team members' prior experience in leading PEFA based assessments and their local country knowledge.

107. The total cost of an assessment for a typical central-government assessment in a medium-sized country with limited input from a government team would be in the order of $100-150,000 including fees, travel cost and subsistence. Much higher costs must be anticipated if the assessment is supposed to cover general government.

108. For very small countries, it may be hard to accept assessment expenses of this magnitude. Clustering of assessments to achieve economies of scale has been discussed for small island states and territories in both the Pacific and the Caribbean, but has proven difficult to arrange as it easily leads to conflict with the principle of active government interest and participation if not leadership.

109. The assessments reviewed reveal that internal organization of the assessment team is not always clearly spelled out in the terms of reference and that the team leader often does not have the means to effectively manage the team (in several cases he/she did not even have access to the terms of reference for the other team members and did not know the exact number of days they had been allocated). This is particularly a problem where a team is composed of members appointed (and financed) by different agencies, whether these members are consultants or agency officials. While this is by no means a unique issue for PEFA based PFM assessments, it is important to highlight that it can easily lead (and in fact has led) to ineffective use of team resources, imbalance in the coverage of the report and consequences for the quality of the final output.

4.7 Sources of information

110. Previous analytical work has typically been used as an initial source of information for the assessors, where it has existed. Only in one case (Ghana) did such material constitute the only source. Assessors have had to depend to varying degrees on such data sources, depending on the number of days allocated for their field work and the extent to which client governments have been helpful in timely provision of the required information.

111. Data collection has been a challenge for many assessments and has absorbed a large part of the assessment teams' field mission time. Rarely has comprehensive data been gathered in advance of the field missions and provided to the team members for preparation and gap identification. The pre-determined resource allocation, therefore, has often constrained the degree of detailed information the assessment teams have been able to obtain and verify during field missions.

112. Information from primary sources has almost exclusively included information from interviews with officials from the assessed government and its agencies. Views of non-government actors (private sector civil society organizations) have rarely been consistently included. Tanzania is a distinct exception, where a civil society representative participated as an assessment team member.

4.8 Quality assurance arrangements

113. Typically three types of arrangements have been in place for quality assurance of the PFM assessments (ref. Annex 1.d) as follows:

- All World Bank led applications have been (or will be when the report reaches the appropriate stage) subject to the Bank's institutional/t review system, based on peer reviewers (internal or external) selected at the Concept Note stage. The peer reviewers (which may include the PEFA Secretariat) contribute both to the Concept Note and to the review of the draft report. The institutional/management review process considers and incorporates the review comments into the product, as appropriate, before it is delivered to the client. Ex-post follow up of this process is provided through the Quality Assurance Group reviews to ensure that team leaders consider and incorporate to a reasonable extent the comments made by the peer reviewers.
- In-country assessment reference groups (typically including the government and donors beyond those directly participating in the assessment team) provide an important review mechanism in that they should have the best impression of the quality and completeness of the information provided in the reports and the extent to which appropriate professional judgment has been exercised in areas of limited data availability. Such arrangements have been made in Kyrgyz, Moldova, Uganda and Mozambique. A weakness of a very comprehensive team of donor agency staff is that few, if any, informed donor representatives will be left in-country to take an arms-length view of the quality of the assessment (Tanzania).
- The PEFA Secretariat has provided a peer review (in World Bank context) or a similar quality review on demand for 11 of the 19 assessment reports. The Secretariat's contribution is mainly to offer an opinion on the extent to which the Framework's assessment scope, methods and report provisions have been met. The Secretariat usually cannot judge the accuracy and completeness of the data used for the exercise, but does assess adequacy of data presented to justify the scoring.

Box IV – Good Practice Examples
Recording diverging opinions in the PFM-PR – Ghana 2006

To fulfill its purpose as common information pool for the Strengthened Approach, a PFM Performance Report should provide a snapshot of PFM performance that all major stakeholders in the assessed country can make use of. While the ideal is that all stakeholders agree on the indicator ratings and conclusions, there are likely to be areas in which opinions divert, particularly between the government on the one side and the donor group on the other. In such cases, the opinions of both parties may be recorded, with details on the nature of differences. The Ghana External Review of PFM (ERPFM) assessment in 2006 (ongoing at end of March 2006, but not yet fully reported) provided the first case of a PFM-PR where such differences are recorded. This is done in footnotes for a few indicators e.g. for PI-8 dimension (ii) as follows:

"The ERPFM team and MoFEP did not reach agreement on this score – the difference related to the timeliness of reliable information available to SN Governments, with the team's assessment that MMDAs did not rely on information prior to the approval of the final transfer amounts by Parliament."

In this case the report reflects the donor-led ERPFM team's assessment with the government's disagreement in the footnote. Similarly, a government-led assessment could record the donor group's diverging opinion in a footnote.

114. Provision for effective quality assurance is essential for the acceptability and shared ownership of the results. Quality assurance needs to focus on both the quality of information collected and on the completeness and correct use of this information for scoring the indicators. It is important to agree such arrangements well in advance of the commencement of the assessment work, as such arrangements will have a bearing on the schedule of events and milestones in conducting the assessment and can influence the team's reporting lines and deliverables (and therefore the terms of reference for consultants).

115. Quality/peer review comments have been provided by the PEFA Secretariat on a few draft reports (all being PFM-PRs) where the impact on the subsequent reports could be measured. In these cases the impact of the PEFA review has been very variable. In some cases roughly half of the detailed comments and suggestions provided on indicator scoring and evidence were adequately responded to, while in other reports the response was extremely limited. The impact of comments on the content and structure of the report and its summary has been even more limited. It is important that such an external review is scheduled relatively early in the report preparation process in order to ensure that the team resources have not been exhausted in dealing with earlier comments and therefore are dealt with in a minimalist manner. It is also important that the leading partner of the assessment ensures that comments are adequately responded to before the final report is issued. The latter appears to be a particularly weak link in the assessment process. An example of a systematic approach to handling comments was found in Moldova, ref. Box V.

Box V – Good Practice Example
An effective quality assurance mechanism – Moldova

Review of the draft Performance Report by major local stakeholders - such as government and donor agencies - as well as external experts is a crucial part of a quality assurance system that leads to a generally accepted report. Equally important is a system that demonstrates how the individual comments have been dealt with by the authors. The latter has often not been clear. A very systematic approach to dealing with review comments was developed in connection with the Moldova assessment and has been adopted in other countries subsequently. This involved both comments from the PEFA Secretariat and comments from the government and donor reference group.

In the case of comments from the PEFA Secretariat, the team amended the first draft report on the basis of comments received (from all reviewers). A second draft report was issued and the PEFA Secretariat received it for comment on the responses to the earlier version. This was done in a matrix format, mostly indicator by indicator, showing initial comment and change in the subsequent report. The team inserted responses in the matrix showing the action they would take on the next draft version of the report plus actions required on supplementary information needed from the government. The final version of the report could then be checked against the matrix, following the development of the report from initial comments to final version.
In a number of cases, comments received from the government and the donor group were partly conflicting. The team developed a matrix of all such comments and a meeting was arranged during which the team went through the areas of conflicting observations with the government and main donors in order to reach agreement on each point.

Due to this process, the response to and impact of comments can be easily tracked from early draft to final report and this creates confidence in the final product.

116. The Secretariat has not had consistent access to comments provided by other stakeholders and therefore cannot judge if their comments have been sufficiently responded to by the assessment team. In the few cases where such comments have been available, however, indications are that the assessment teams' responses have been limited also to those comments.

4.9 Integration into broader analytical products versus free-standing PFM-PR

117. Some PFM assessments have benefited from an almost simultaneous undertaking of complementary analytical work, so that economies of scale could be gained in data gathering and use of consultants. Also the ability to link PFM assessments to work on public expenditure reviews has been beneficial. Such simultaneous work took place in Kyrgyz (PPER) and Panama (PER).

118. PEFA-based PFM assessments have been integrated into or implemented under parallel coordinated arrangements with analytical work with a broader scope in several cases, (ref. also section 2.2 and table 3) such as:

- PERs in Afghanistan and Tanzania
- CFAA in Guatemala, Tanzania and Republic of Congo
- detailed procurement assessment (CPAR) in Afghanistan, Guatemala, Panama and Bangladesh
- Fiduciary Risk Assessments in Ghana, Moldova and Guatemala.

119. **Integration with PER, CFAA and CPAR –** There are distinct potential advantages of combining the Framework with broader analytical work. The CFAA elements to a large extent overlap with the scope of most of the indicators of the Framework, but may go deeper in a number of PFM subjects, which would help in providing more substantial evidence for the indicator scoring. Even if revenue administration is not included in CFAA guidelines, it is an area that has been covered in a number of CFAAs as the need to look at the issue emerged e.g. in Latin America. Similarly the CPAR and the use of the detailed OECD-DAC indicators for drill-down on procurement issues could lead to better basis for scoring the PEFA PFM indicators. PER work also covers some of the areas of the PEFA indicators, particularly the indicators of quantitative analysis of budget outturns and issues concerning the scope and formulation processes for the budget. Thereby, the indicator based PFM analysis may be linked more firmly to the budgetary outcomes in the summary assessment (or equivalent section in integrated products). In addition, PER work may provide the information required to link the PFM assessment to the achievement of intended budgetary outcomes (ref. para. 72 above).

120. The experience so far is that the potential linkages have been used mainly in relation to the PER elements of broader analysis. Integration between PEFA indicators and CFAA analysis has been mixed with quite close integration in some reports and addition of the PEFA Framework appearing as an afterthought in others. Procurement assessments are consistently done as largely free-standing work with little indication of linkages to other parts of the analysis.

121. A lower degree of compliance with PEFA principles and methodology of indicator scoring is found when the Framework is integrated directly into broader work than when it is free-standing either as a separate exercise or the subject of a separate report of a dual analytical exercise (ref. section 3.2, table 5). This is by no means to say that overall quality of the entire integrated product is lower, but rather that the indicator scoring may have been given less attention. There is no reason why this could not be improved upon. Particularly interesting is the emergence of ‘dual’ analytical products where the compliance with PEFA methodology is as high as in free-standing PFM-PRs, while at the same time the other part of the product satisfies the demands from other stakeholders in the exercise, such as e.g. detailed diagnostic analysis of select areas and related recommendations for capacity development.

122. The time taken to complete an assessment varies significantly. Free-standing PFM-PRs are generally completed with a final report in 3-6 months from start of field work. The shortest production period is found for Malawi (4 months from issue of terms of reference till issue of final report). Combined analytical work with the PEFA Framework integrated or as a separate report take much longer. Most of these assessments had not been issued in a final report 12 months or more after the completion of the concept note or terms of reference, except where the PFM-PR is combined with an FRA which hardly added to the production time (ref. the diagram in section 2.1, table 2).

123. Long production periods for a performance assessment are a concern for several reasons. The assessment is supposed to provide a snapshot of the performance but long periods of discussion leads to uncertainty as to what moment the snapshot represents. This undermines monitoring of progress. Donor agencies that need the assessment as input to internal fiduciary considerations may be tempted to revert to separate exercises if a final assessment cannot be produced with in reasonable time.

124. The ‘dual product’ option holds promise of a solution. Appropriate packaging and sequencing the different components of the analytical work may allow the PFM-PR part of the work to be completed relatively early in the process, while work continues on other parts e.g. on strategic expenditure allocation and other more policy related issues, or on detailed identification of the underlying causes for performance in selected areas.

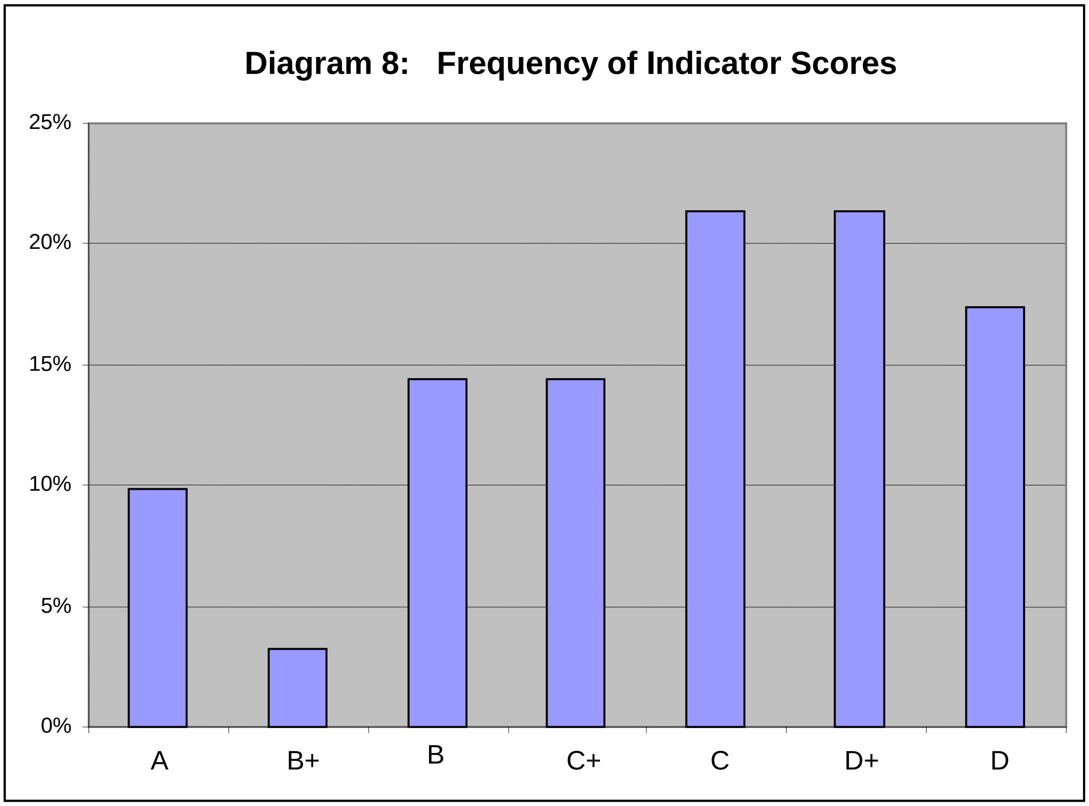
125. **Fiduciary risk assessments** - Where donor agencies need to undertake a fiduciary risk analysis for their grant/lending operations, there are distinct advantages of separating such a FRA from the PFM-PR in different reports, but in such a way that the PFM-PR provides an important input to the FRA. Firstly, the concept of fiduciary risk is not well defined and there is no international agreement on a definition and the operationalization of such an assessment, e.g. the method of aggregation to reach an overall fiduciary risk score. Donor agencies may attach different weights to different aspects of the PFM assessment in their FRA and add elements that are not covered directly by the PFM-PR, such as level of corruption and anti-corruption measures. Secondly, while a client government and the country donor group may be able to agree on (or accept) the scoring of the PEFA indicators and a narrative summary based thereon, the FRA results may be much more controversial. This could make it difficult to use the report for the ongoing dialogue with government on reform formulation and monitoring. Also, the government may not wish a fiduciary risk rating by the donors to become disclosed to third parties as it may affect commercial credit ratings (a concern raised by officials working in ECA and LAC regions). This could lead to very restricted circulation of the report and defeat other potential usage of the PFM-PR such as discussion in Parliament.

4.10 Managing expectations

126. The Consultative Draft on the Strengthened Approach recommends that PFM assessments on the basis of the Framework be undertaken at intervals typically in the order of 3-5 years due to the moderate rate of progress in PFM performance that would be expected in most countries and therefore the likelihood that only very minor changes in improvement would be noticeable in assessments based on high-level indicators with only 1-2 year intervals.

127. During the introduction phase of the Framework in any given country, it will be necessary to manage the government’s (and donors’) expectations as regards the anticipated levels of performance ratings. Diagram 8 shows the distribution of indicator scores as presented in the 19 country assessment reports representing all regions and a mixture of least developed, low income and middle income countries. The median score is a ‘C’, which – together with a ‘D+’ - is also the most frequently assigned score. It should be noted that while the Secretariat’s reviewers sometimes concluded differently on the individual indicator ratings, the overall distribution of the reviewers’ scoring follows practically the same distribution

.



128. As mentioned earlier, comparison of scores by region, size, government structure, income level etc is not yet possible on a firm basis. However, it would be fair to expect that a typical low-income country would see the majority of its indicator scores in the C-D range with only exceptional indicators scoring a ‘B’ or an ‘A’. It would be reasonable to sensitize a government to this reality before embarking on a first application of the Framework in a given country.

129. In HIPC AAP countries in particular, it would be important to emphasize that a particular letter-score in the PEFA Framework’s scale represents a different performance level from the same letter-score in the HIPC AAP framework, which only used 3-point ordinal scale of A, B, C. It is also important to emphasize that there are no standardized benchmarks to be met in the PEFA Framework, in contrast to the HIPC AAP framework. The use of similar scoring nomenclature but with different meaning has led to misunderstandings in a few cases.

130. In addition to expectations regarding indicator scoring, it is important to manage expectations as to what a PEFA based assessment can contribute and not contribute in relation to PFM reform formulation and monitoring. A PEFA based assessment may assist the government and it’s supporting donors to identify weaknesses in PFM performance that should be addressed as a priority. While the PEFA Framework can help to identify areas where detailed work is needed, the Framework is not a detailed diagnostic instrument, designed to offer detailed recommendations for capacity building measures. Moreover, technical linkages of the PFM system development and the realities of political economy for reform are equally important ingredients in determining a reform strategy and are not discussed in the PFM Performance Report, which therefore does not contain all the element required as a basis for formulation of a PFM reform program and action plan.

4.11 Dissemination of reports

131. Seven of the reports reviewed had reached the final version by end of March 2005. The World Bank has established procedures for agreeing with the client government on public access to a report, e.g. through posting on the Bank's external website. PEFA based PFM assessments for Afghanistan, Tanzania, and Bangladesh (the latter in the Country Assistance Strategy), have been made available to the public through this means.

132. No prior agreement or specific procedures for agreement on dissemination of final reports appears to have been made for assessments led by other donor agencies. These reports are shared between the government and the donors represented in the country, but plans for, and acceptance of, any wider dissemination are unclear. The terms of reference or concept note do not discuss the dissemination of the reports after completion. Dissemination is typically left for discussion till after the assessment report has been virtually completed. In practice, informal sharing of the reports among donor agencies is quite widespread. A few of these reports have eventually emerged on either government websites (Mozambique) or donor agency websites (the Zambia report on DFID's website, Kyrgyz Repulic on the World Bank's PEMPAL website).

4.12 Future update of the assessment

133. For most countries, the 19 reviewed assessments represent the first PFM assessment using the Framework. In some of these countries, however, testing of the draft 2004 version of the Framework was undertaken, but not with government participation and never taken forward to a completed assessment report, as the purpose was limited to learning on the Framework content. It has therefore been impossible to make meaningful comparison of the results. Up-front agreements on when and how the PFM assessment should be repeated are rare, but there are some examples such as Zambia (PEFA Assessments scheduled for 2008 and 2010 in relation to the planned implementation of the government's PFM reform program).

134. A number of reports have pointed out the difficulty of improving scores over the short term of about one year.[8] This contrasts with the need of governments and donors for more frequent review, not only for monitoring reform activities but also for evaluating their high-level impacts and thus informing donors' budget support decisions. For instance, direct budget support agreements often require annual performance reviews. The report on Mozambique recommended that full PEFA reviews be undertaken at three year intervals and that, in the intervening years, a 'reduced assessment' be made by a local joint team.

[8] This has given rise to the use of arrows where the score has not changed but improvements have been made, and it is important to recognize them and give credit to governments.

135. PFM performance assessments have been planned for 2006 in some of the countries included in this review (Malawi, Tanzania and Ghana), all of which were ongoing at the time of finalizing this report on early experience. The reasons for such an early ‘repeat’ are different. In the case of Malawi the 2005 assessment was directly linked to decisions on direct budget support programs. The repeat assessment in 2006 will be used as a platform for discussing revision of the reform program and involve more broadly the donor community. In Tanzania, the framework has been incorporated in the annual PER external assessment, which serves a set of purposes including provision of information for budget support decisions and monitoring PFM reforms. In Ghana, the 2005 applications mainly served internal donor fiduciary purposes and did not or to a limited degree involve the government and the local donor community at large. In this case, therefore, the 2005 assessment did not serve the purpose of a ‘common information pool’ agreed to by all major stakeholders. The 2006 assessment will be considered the baseline assessments for future tracking of performance change and reform progress.

4.13 Conclusions regarding the process of planning and implementing an assessment

136. A distinction was initially expected to arise between ‘independent’, ‘external’, ‘joint government-donor’, and ‘government self-‘ assessment in consideration of the quality and objectivity of the results. The experience shows, however, that such differences are minimal and that the best solution may be one that includes strong government involvement, multi-donor participation in planning and review as well as inputs from external or independent actors such as experienced consultants.

137. Government involvement in the process is crucial for ensuring the best possible information basis for the assessment and subsequent use of the assessment for reform impact monitoring and dialogue on reform priorities. Government ownership of the assessment is facilitated where early orientation, training and team building is undertaken through a joint government-donor workshop, where the government designates its own counterpart team to facilitate and participate in the assessment and where joint discussion of assessment results and their implications are thoroughly discussed e.g. in a restitution workshop.

138. Managing expectations is important in relation to PFM assessment in any country. This concerns (i) the likely level of ratings, which for most low income countries would be predominantly in the C to D range, (ii) changes from earlier assessment that are likely to be modest within a 1-2 year horizon even in actively reforming countries. But expectations should also be realistic as to (iii) what the PFM performance assessment can contribute to and what it cannot do.

139. The resources required to undertake a PEFA based PFM assessment has often been under-estimated. This has led to some cases where inadequate information was available to support the assessment and other cases where additional resource inputs had to be allocated mid-way through the assessment. A number of assessments have counted on a single, generalist PFM expert to undertake assessments, which has led to some PFM areas being inadequately covered. A team of two experienced assessors with complementary background, supplemented as needed by short inputs from a few specialists, appears to be a suitable formula, especially if combined with a local consultant or a government counterpart team. A total team input of 3-5 person-months should be foreseen for a central government assessment in a typical medium-sized country if implemented as a stand-alone PFM-PR.

140. PEFA based PFM assessments are often combined with broader analytical work such as PER, CFAA, CPAR and FRA (or combinations of them). This can lead to tensions between different stakeholder interests in terms of the role of the performance assessment, the time taken to finalize it, the adherence to the PEFA Framework and the links to recommendations for reform formulation and action planning. 'Dual' analytical products that prepare a separate PFM performance report as part of or in a coordinated parallel process with the broader analytical work offer the prospect of embracing all stakeholder interests. With appropriate packaging and sequencing of the components, such a 'dual product' could become an important instrument in furthering all the elements of the Strengthened Approach.

141. A well-defined and well-managed quality assurance mechanism is crucial for achieving a final assessment of high quality and acceptable to all stakeholders. Such a mechanism should be agreed among all stakeholders at en early stage of planning for the assessment. The agency/person responsible for managing the mechanism should be identified and needs to systematically keep track of comments and how the report authors respond to them and reflect them in subsequent report versions. A combination of a local reference group of donor agencies and government and external reviewers from agency headquarters and/or the PEFA Secretariat appears to secure the best mix of views in terms of data reliability and compliance with the PEFA principles and methodology.

142. In summary, the application of the PEFA Framework for PFM performance assessment is a very decentralized process with the government and the locally involved donor agency group at the centre of planning and managing the assessment work. These stakeholders need in particular to consider and agree - ideally at an early stage in the planning process - on the following issues:

- packaging and sequencing of diagnostic products / reporting of the indicators (e.g., standalone PFM-PR, "dual products", or integration of the indicators into other products and the extent of any recommendations in any component of such a package);
- the role of various parties in conducting the assessments (ranging from government self-assessments with external validation, to assessments that are principally done by external partners with government collaboration, the leading manager of the exercise)
- the related financial and personnel resources required;
- the time for finalization of reports (taking into account the need to ensure client understanding and ownership);
- quality assurance arrangements; and
- report disclosure arrangements.

143. There are indications that the Framework has facilitated donor harmonization and collaboration around PFM analytical work, but it is too early to judge to which extent the Framework has reduced the number of PFM diagnostics and promoted donor coordination as concerns reform dialogue with government and coordination of technical and financial support.

5 Application of the PEFA Framework at Sub-National government level

5.1 Extent of PEFA based assessments of SN government performance

144. This chapter provides an overview of where, why and how the Framework has been applied to PFM assessments for Sub-National (SN) Governments. The main focus of the analysis is to examine the ease and suitability with which the Framework has been adaptable to SN level.

145. During 2005 the Framework has been used for PFM assessments at SN level in four cases. Annex 5 provides key data on those four cases, in Ethiopia, Pakistan (Punjab State only), Mexico and Uganda. Only the Uganda application has used the June 2005 version of the Framework as its basis; the other three refer to the 2004 version as they were all well under way before even the first complete draft of the 2005 version was (unofficially) available in April 2005. The Uganda case is, therefore, the only one that can lead to detailed conclusions and lessons on how to adapt the final Framework to the SN level. The other three cases, however, can contribute general lessons on use of PEFA based methodology at SN level. All four cases are distinctively different, but apart from Mexico, the assessments were instigated and managed by donor agencies. An assessment similar to the one in Uganda is under way in Tanzania. In the North West Frontier Province of Pakistan, an assessment is ongoing, similar to the one in Punjab, but using the 2005 version of the Framework. In Nigeria, an assessment is ongoing that seeks to combined a PFM assessment of central government with assessments of three regional state governments.

5.2 Country cases

146. In **Mexico** the Framework has been used as a basis for developing a system of self-assessment for regional states by voluntary self-selection and is supported by a technical assistance operation by the World Bank and the Inter-American Development Bank. Mexico is developing an indicator set that is different from, but inspired by and mapped against, the PEFA indicator set. It is specifically designed for publication on the respective state governments' websites in order to present state of and progress in PFM in order to promote private investment and possibly supplement the commercial credit ratings.

147. The **Pakistan** case concerns only one regional state, Punjab Province. However, Punjab has a population of more than 70 million and is therefore larger than practically all the countries for which a central government PFM assessment has been undertaken. The assessment for such a regional state is very similar to the typical assessment of central government PFM. The reason for this must be sought both in the size of the regional state and in the provision of donor funding specifically for the regional state (even if central government approval and debt service guarantee is required), which calls for both a fiduciary assessment and for tracking the impact of PFM reform activities. The Punjab assessment made use of the full set of PEFA indicators and PEFA scoring methodology (but the 2004 draft version, as earlier mentioned).

148. In **Ethiopia**, a sub-set of 14 PEFA indicators (2004 version) was used, but not the PEFA scoring methodology and calibration. Scoring was done on a three point ordinal scale comparable to HIPC AAP and DFID's fiduciary risk assessment methodology. The Ethiopia assessment seeks

to arrive at an overall assessment of the PFM system for general government (i.e. central and sub-national). The driver for this approach has been the provision of general budget support to the central government, which in turn transfers substantial amounts of grants to sub-national entities outside central government's direct control. The same set of indicators has been applied to central government and a sample of the top layer of sub-national governments. In Ethiopia, the sample comprised of six SN governments (out of a total of 10), representing 98% of the national population.

149. The **Uganda** assessment also attempts to arrive at an overall assessment of the PFM system for all of general government (i.e. central and sub-national). The justification for this expansion of the scope of the PFM performance assessment is that some 35% of central government expenditure constitutes transfers to sub-national governments at municipality and district level. In Uganda, the sample also comprised six upper-tier SN governments (out of 64 districts and 13 municipalities), but they represent only 10% of the national population. The SN level application in Uganda, therefore, was the only one that faced problems with determining the size and criteria for selecting a sample and with aggregation of results for the SN governments to reach national level conclusions for SN government itself, or when combined with the central government assessment, for the general government sector. The assessors assumed on the basis of similar analytical work in 2004, that the problems at district level were generic so that aggregation problems would be minimal. However, the draft report does not explain how the aggregation of the ratings for the sample of the six SN governments to national level conclusions for SN government performance is done. It is also uncertain if/how the central government and SN government performance assessments are going to be combined into an assessment of PFM performance for the general government sector. The final report may yet clarify those issues.

150. Uganda made use of the full set of PEFA indicators and PEFA scoring methodology and applied similar sets of indicators (with appropriate amendments at SN level) to both central and SN government assessments. Only the Uganda LG assessment offers opportunities for learning about the modifications needed to adapt the final version of the PEFA performance indicator set to SN government circumstances and any problems encountered in the field application. The experience here is that

- The modifications to the indicator set and design of five new tailor-made indicators was left entirely to the consultants. They were very loyal to the PEFA methodology and the changes made appear generally reasonable, with one exception concerning an indicator which is policy based and therefore impossible to calibrate on the basis of good PFM practice (degree of independence of financial management at district government level).
- While approved by the donor group, very little effort was contributed from government and donors in ensuring that the modifications and additions were appropriate. Such efforts could have improved the standard of the modifications and additions, but very little time was available before the start of the field work for such consultations.

151. The compliance rates for indicator scoring is significantly lower for the local government assessment than for the central government assessment. Indications are that the resources allocated to undertake a full and proper assessment in each of the sample SN governments was under-estimated. This resource use is estimated at 4 consulting months for the internal consultant and 4 local consultants who carried out the work, organized in groups of 2-3 consultants doing the field

work in each district or municipality. This may seem a very substantial increase in the resources for adding the SN government assessment to that of central government (more than 200%) but it constitutes only 2-3 weeks consulting input for each SN government. While there will be some economies of scale in assessing six SN governments with similar rules and structures in one exercise, one should not forget that data for each indicator has to be collected and judged for each of the SN government and that each of these SN government administration has the same size and financial turnover as many small states. As mentioned in chapter 4, a PFM assessment of a government of a small state is only slightly smaller than the resources needed to assess a medium size state

5.3 Conclusions regarding application at SN government level

152. On the basis of the so far limited use of the Framework at SN government level, one can tentatively conclude, that the use of the Framework is feasible for individual SN governments with minor modification of the standard indicator set and addition of a few indicators to reflect the special circumstances of SN governments. However, no model is available for selection and aggregation of results of a sample of SN governments to reach conclusions at the national level and for general government.

153. The resource requirements needed for a performance assessment of a large regional state may be the same as for a medium sized national government. However, undertaking a full performance assessment for SN government, even on a sample basis in a medium sized country, can add a very substantial resource demand to a central government assessment. There is a clear risk that such a resource demand is under-estimated and as a result the compliance with assessment methodology and quality of reporting will be below the standard for the central government assessments.

154. Final conclusions on the methodology and resources used will not be possible until full quality review of the report has been completed including the government response. Such conclusions will also be more robust when experience from the ongoing work in Tanzania and Pakistan/NWF Province and Nigeria can be incorporated.

6. Recommendations

155. **All partners** (including PEFA partners, other donor agencies and partner governments) **should focus on measures to improve quality of the assessments as the first priority.** Compliance with the principles and methodology of the PEFA Framework is key to transparency and consistency in rating and therefore to confidence in and use of the Framework.

156. **Donor agencies need to ensure that both head office and country level staff are well acquainted with the Framework as least for assessment planning and management purposes.** This concerns both PEFA partners and non-PEFA partners**.** The objectives of the Strengthened Approach; the role, benefits and limitations of the PFM Performance Measurement Framework; and the importance of government participation, broad donor participation, adequate assessment team resourcing and the quality review mechanism for the PFM assessment are key points to bring across to most donor agency staff.

157. **The benefits of the Strengthened Approach and the PFM Performance Measurement Framework should be brought across to partner governments in a much more systematic fashion that hitherto.** Government officials at executive level should be informed of the potential benefits the approach and the tool may offer them in terms of donor harmonization and reform impact monitoring. Officials at management level should be informed about the principles and methodology of the Framework in order to effectively participate in and contribute to the assessment work.

158. **Assessors that constitute the PFM assessment teams and performance report authors need to be consistently briefed before and supported during the assessment work.** Review of completed reports have provided many clues to where difficulties arise in information collection, indicator interpretation and summarizing of assessment.

159. The PEFA Program should support the above aims by **developing a dissemination and training strategy and program** that effectively assists donor agencies, partner government and consultants along these lines;

160. In terms of technical work, t**he PEFA Program should develop further clarification and guidance** on a continuous basis, including

(a) clarification on the role of the framework and on its limitations;

(b) clarifications to the interpretation of the indicators, especially those that have proven most difficult to assess (such as indicators PI-24 and PI-25);

(c) guidance on the evidence that would be adequate for scoring the indicators and likely sources of such evidence with emphasis on methods of collecting evidence for indicators where this has been a specific problem and on triangulating with information from non-government sources (e.g. for indicators PI-7 and D-1);

(d) guidance on planning and management of assessments, drawing on the examples of good practice identified so far (a first attempt in terms of checklist for assessment planning is included in Annex 5)

Such clarification and guidance should be made available of the website and incorporated into training programs and materials

161. **This review should be repeated in mid 2007.** A further year of experience would in particular allow more experience to be gathered on the use of the PFM performance assessments and their contributions to furthering the Strengthened Approach as well as experience from a few cases of repeat assessments.

162. During this period the PEFA Program should also monitor utilization of the Framework at sub-national government and sector level as well as cases where non-tax revenue is an essential part of total government revenue, drawing lessons and issuing guidance in this respect.

.

ANNEXES

Annex 1.a List of Country Assessments considered for this Review[9]

	Country	Donor Agencies involved (**lead agency** + others)	Most recent report version considered for this Review
Sub-Saharan Africa	Congo Brazzaville	**WB** (overall incl. CFAA), **EC** (PFM-PR)	PFM performance assessment. Version Provisoire, March 2006
	Ghana	**DFID**, DK, Switzerland, EC	PFM Performance Assessment Report, Desk draft 2.2, Sept 28, 2005
	Malawi	**EC** with DFID, Germany, Norway/Sweden	PFM Assessment Final Report July 18, 2005
	Mozambique	**EC** with 16 other donors	Assessment of PFM in Mozambique 2004/05. Final report, March 2006
	Lesotho	**WB**	Public Expenditure and Financial Accountability Review, Review of PFM, February 2006
	Tanzania	**WB** with IMF, EC DFID, Norway, Switzerland, Germany, AfDB, Canada	Public Expenditure & Financial Accountability Review FY05, Draft final Oct 2005
	Uganda	**EC,** WB, DFID, Norway, Ireland and Sweden, Germany, AfDB, Japan	PFM Performance Report and Update of CIFA Action Plan, Draft Nov 2005; Local Government PFM Assessment, Draft Dec 21 2005;
	Zambia	**Sweden**, DFID, WB, EC, Netherlands, Ireland, Denmark, Finland, Germany, Norway, UN	PFM Performance Report and Performance Indicators, Final report Dec 2005
Latin America & Caribbean	Guatemala	**WB, IADB**	Country Financial Accountability and Procurement Assessment Report, Draft June 24, 2005
	Panama	**WB, IADB**	Country Financial Accountability and Procurement Assessment Report, Confidential Draft March 31, 2006
South Asia	Afghanistan	**WB**, IMF coordinated with DFID, EC and AsDB	Managing Public Finances for Development, Final Oct 6, 2005
	Bangladesh	**WB**, DFID, Japan, AsDB	WB Country Assistance Strategy Feb.2 2006 based on Review of Institutional Arrangements for Public Expenditure Financial Management & Procurement. Draft July 19, 2005
Europe & Central Asia	Moldova	**EC,** Sweden, UNDP, WB, IMF	PFM Performance Report, Draft, February 19, 2006
	Kyrgyz Republic	**DFID, Switzerland**, WB, Sweden with IMF, EC	PEFA PFM Assessment, Final Report, January 18, 2006

[9] The data in Annex 1.a, 1.b, 1.c and 1.d includes the 17 reports that were or will be completed with substantial use of the PEFA Framework, but excludes the two reports where the Framework was initially included in a non-substantial fashion and which were later abandoned.

East Asia & Pacific	Fiji	**WB,** Australia	PFM Performance Report & Performance Indicators. June 2005
	Papua New Guinea	**WB**, AsDB, Australia	PFM Performance Report & Performance Indicators. September 2005
Middle East & North Africa	Syria	**IMF,** WB	Public Finance Management (TA report) draft March 2006

Annex 1.b Assessment Purpose, Roles and Resources

Country	Purpose of the assessment	Government's role	Assessors	Personnel use
Congo Brazzaville	Assessment of PFM performance to inform identification of reform program and to set baseline for monitoring reform impact	Provision of information	Consultant (supported by WB CFAA team)	1 internatl (2 PM) + input from agency staff
Ghana	Assessment of fiduciary risk for bilateral budget support	None	Multi-agency team w/consultant	7 agency staff + 1 intl consultant
Malawi	PFM assessment was condition for disbursement of EC GBS grant	Passive, but very helpful in providing access to data and officials for interview	Consultants	2 internatl (2 PM)
Mozambique	Set benchmarks for tracking future progress in PFM reform; review and adjust present reform plans; meet donor needs for GBS	Actively involved from initial workshop, through data collection and discussion of implications for reform	Consultants	2 internatl + 1 local (5 PM)
Lesotho	Satisfy WB knowledge mandate; feed into dialogue with govt on future GBS operations.	Provision of information	Single-agency team w/consultant	1 internatl (1 PM) + 2 agency staff
Tanzania	Provide Govt & DPs with a comprehensive, integrated and candid assessment key fiduciary risks and recommend improvements to the PFM framework, institutional performance & capacity building	Provision of data, commenting on report	Multi-agency team w/consultants	17 agency staff, 4 intl consultants, 1 NGO repres. (of which 4 members were directly involved in the indicator ratings)
Uganda	Assess the current status of PFM; establish progress in implementation of the CIFA Action Plan of 2004 and recommend amendments	Provision of data; comments on report (not yet made)	Consultants	CG: 2 internatl (1.5 PM) LG: 1 intl + 4 local (4 PM)
Zambia	Establish baseline for monitoring progress on PEMFA program implementation and recommend sequencing of its 12 components	Assessment led by government team	Government team w/consultants	5 govt officials + 2 internatl consultant (2.5+1.5 PM)
Guatemala	Provide an updated assessment of PFM practices incl. Procurement as basis for a reform program.	Provision of information. Comments on draft report. Participated in workshop on action plan.	Two-agency team w/consultant	8 agency staff + 1 intl consultant

Country	Purpose of the assessment	Government's role	Assessors	Personnel use
Panama	Provide an updated assessment of PFM practices incl. Procurement as basis for a reform program.	Provision of info. Draft is subject to govt comments.	Two-agency team w/consultant	7 agency staff + 1 intl consultant
Afghanistan	Assess progress since end of Taliban era and present status of PFM. To set out a road map for PFM reform.	?	Multi-agency team	4 agency staff
Bangladesh	Stocktaking of PFM arrangements as basis for agreeing future reforms; forming part of CAS results monitoring framework	Provision of information; discussion and agreement on baseline scores	Two-agency team with government validation	?
Moldova	(i) provide a common donor platform for assessing fiduciary risk and (ii) serve as a vehicle to the government and donors for tracking PFM performance progress over time	Provision of data and interviews. Discussion of findings after 1st consulting mission. Formation of PEFA team to review indicator assessments in draft report.	Consultants	3 internatl + 2 local
Kyrgyz Republic	Establish baseline for planning & monitoring PFM reform; input to fiduciary risk analysis toDFID and SECO for GBS programs	Chaired a PEFA working group. Participated in pre-mission briefings. Provided information. Discussed findings and ratings.	Consultants w/some agency input	2 internatl (2 PM) + 3 agency staff
Fiji	Review of the PFM component of the Public Sector Reform Program	Formed team of fficials to work with consultant on data collection and ratings	Consultant w/government team	1 internatl (1 PM) + 4 govt officials
Papua New Guinea	Part of public sector reform program review; to support public expenditure review and rationalization.	Provision of data/ information	Consultant w/ support by multi-agency team	1 internatl (1 PM) + ? agency input
Syria	Establish baseline for planning & monitoring PFM reform as part of technical assistance report on identifying and prioritizing reform actions	Provision of data/ information	Two-agency team w/consultant	7 agency staff + 1 intl consultant

PM = person month

Resource use excludes planning, administration, head office management and report reviews

Annex 1.c Report Characteristics

Country	Nature of assessment and report	**Length** (main, summary, annexes)	Inclusion of Recommendations	Report sharing as at March 31 2006
Congo Brazzaville	PFM-PR as a separate report coordinated with a CFAA	107 pages (93+4+5)	None	Internal to EC & WB only
Ghana	Standard PFM-PR; Separate report as part of DFID Ghana Fiduciary Risk Assessment, Final Oct 20, 2005. Desk assessment.	50 pages (36+5+9)	None	?
Malawi	Standard PFM-PR	102 pages (61+11+ 30)	None	Government, CABS group, WB, IMF
Mozambique	PFM-PR.	89 pages (46+5+31)	Reform recommendations and performance outlook for 2006	Government, 16 donor agencies,
Lesotho	Desk study with short field mission; FM input to PEFAR report; 1st draft, internal working document	37 pages (34+2+1)	Recommendations made for each main area	WB internal
Tanzania	Integrated PER external evaluation and CFAA/PFM assessment incorporating PEFA indicator assessment as 14 page annex	189 pages (107+15+67)	Recommendations made for each main area	Government and all donor agencies
Uganda	Standard PFM-PR for central government. A separate report for local government based on simple PFM-PRs for each of 6 LGAs.	51 pages (30+3+18) for central govt	Annexes on reform action plan (CIFA) implementation and amendments	Government + donor group
Zambia	Standard PFM-PR	83 pages (39+5+39)	Recommendations for reform planning and sequencing.	Government and donor group
Guatemala	Integrated CFAA/CPAR incorporating PEFA indicators, feeding into fiduciary risk assessment as separate note	130 pages (63+5+55)	Recommmendations for all main areas. Feeds into action plan agreed with government for final report.	Government + WB+IADB
Panama	Integrated CFAA/CPAR incorporating PEFA indicators, feeding into fiduciary risk assessment as separate note	Pages 196 (58+7+52) + 80 page procuremt annex	Recommmendations for all main areas. To feed into action plan after response from govt.	WB+IADB

Country	Nature of assessment and report	**Length** (main, summary, annexes)	Inclusion of Recommendations	Report sharing as at March 31 2006
Afghanistan	Integrated PER/CFAA/CPAR product with PFM-PR incorporated as Vol.II. Part 1	48 pages (37+5+6)	Road map for PFM reform with short and medium term priorities	Published as WB grey cover report
Bangladesh	Part of a PER policy note to be completed in 2006. Incorporates detailed procurement assessment and IPSAS gap analysis plus summary indicator table in Country Assistance Strategy	42 pages (31+4+4) + 22 page procuremt 8 pages in CAS	Agenda for PFM reform (3-page table of recommendations)	Government + WB+ DFID + Japan + AsDB
Moldova	Standard PFM-PR. feeding into separate Fiduciary Risk Assessment, done simultaneously.	56 pages (45+7+4)	A few in connection with specific indicators (PI-24 and PI-25).	Government + Moldova donor group
Kyrgyz Republic	Standard PFM PR	45 pages (34+4+7)	None	Government + donor group
Fiji	Standard PFM-PR.	44 pages (31+4+9)	Includes recommendations for reform program	Government + participating donors
Papua New Guinea	Standard PFM-PR.	47 pages (34+4+9)	Broad recommendations on reform management and prioritization	Government + participating donors
Syria	IMF technical assistance report format with indicator ratings table as an annex.	94 pages (61+13+13)	Yes, the main purpose	IMF+WB

Annex 1.d Indicators Use, Methodology and Quality Assurance

Country	Indicators used	Scoring methodology	Quality Assurance arrangements
Congo Brazzaville	28+3 PEFA indicators, but PI-28 not scored due to not being applicable	General compliance with PEFA Methodology. No arrows used.	WB, EC and PEFA review
Ghana	28+3 PEFA indicators, but 9 indicators not scored for lack of data (desk study)	General compliance with PEFA Methodology incl. upward arrow.	PEFA review
Malawi	28+3 PEFA indicators	Basic compliance with PEFA Methodology. Arrows in three directions used to indicate direction of change.	Comments from country GBS group and govt.
Mozambique	28+3 PEFA indicators.	General compliance with PEFA Methodology. No arrows used. Includes a set of forecast scores two years on.	Comments from country GBS group and govt, and PEFA review
Lesotho	20+2 PEFA indicators (selected as per CFAA coverage)	Dimensions not scored; No arrows used	None so far (WB peer review for complete PEFAR)
Tanzania	28+3 PEFA indicators + 1 additional corruption indicator	Indicator dimensions not scored. Some indicators scored partly against 2004 calibration. No arrows used.	WB peer review incl. PEFA
Uganda	28+3 PEFA indicators, incorporates LG PFM Assessment using the same indicators	General compliance with PEFA Methodology. No arrows used.	Comments from country GBS group and govt, but govt not yet commented.
Zambia	28+3 PEFA indicators.	Basic compliance with PEFA Methodology. Score D used as default for lack of info. Upward arrow used twice (incorrectly).	Comments from country GBS group and govt, and PEFA review.
Guatemala	26 PEFA indicators (excl. Donor indicators and PI-18, PI-19)	Basically using PEFA methodology by highlighting status in calibration text. No arrows used.	WB peer review
Panama	26 PEFA indicators (excl. PI-8 for lack of relevance, PI-23 and donor indictors)	Basically using PEFA methodology by highlighting status in calibration text. No arrows used.	WB peer review incl. PEFA
Afghanistan	28+3 PEFA indicators	General compliance with PEFA Methodology. Numbers 4,3,2,1 used instead of A-D. No arrows used.	WB peer review incl. PEFA

Country	Indicators used	Scoring methodology	Quality Assurance arrangements
Bangladesh	CAS uses 28+3 PEFA indicators of June 2005 (re-worked from the July policy note which used 36 indicators, mostly picked from PEFA indicator set, both 2004 draft and 2005 versions).	CAS based on PEFA methodology but not showing dimensional scores and only showing summary of justification. (July policy note did not use PEFA methodology. Universal calibration used for indicators).	WB peer review
Moldova	28+3 PEFA indicators (D-1 not scored due to irrelevance)	General compliance with PEFA Methodology. No arrows used. D score used as default for lack of information.	Review by members of donor group, government and PEFA review.
Kyrgyz Republic	28+3 PEFA indicators (6 not scored due to lack of data)	General compliance with PEFA Methodology. Upward arrows used once. D score used as default for lack of information.	Comments from country GBS group, WB and IMF and govt, and PEFA review.
Fiji	28+3 PEFA indicators	General compliance with PEFA Methodology. No arrows used. Some confusion on use of '+'.	PEFA review
Papua New Guinea	28+3 PEFA indicators (D-1 not scored due to irrelevance)	General compliance with PEFA Methodology. Upward arrows used once (incorrectly)	?
Syria	24+0 PEFA indicators, due to defined scope of study	General compliance with PEFA Methodology. Upward arrows not used	IMF + WB peer review

Annex 2

Measuring Compliance with Scoring Methodology

Compliance Index

1. In order to assess the compliance with indicator scoring a simple compliance index has been developed, which reflects the adequacy of evidence presented for the scoring and the correct use of such evidence to arrive at a score. These are the two fundamental elements in scoring any indicator. The aggregated assessments of these elements are then related to the total number of scores presented. Where no score has been presented, the indicator is neither included in the numerator nor in the denominator. Thus the quality index becomes neutral to un-scored indicators. The neutrality is introduced to avoid a bias in the compliance index potentially caused by two situations. On the one hand, it is good practice not to score an indicator if evidence is insufficient. Non-scoring in such circumstances should not result in a lower quality index. On the other hand, there are reasons of poor practice that may lead to an indicator not being scored (e.g. inadequate allocation of resources for the assessment) and non-scoring in those cases should not lead to a higher compliance index. Again, the defined scope of the assessment may lead to some indicators not being covered, which in most cases should be considered a neutral reason. However, it is not always clear what is the actual reason for not scoring an indicator. In summary:

2. Compliance Index = (Number of scores with sufficient evidence presented AND correct use of the evidence for determining the score) divided by (number of scores)

3. The compliance index is calculated for

- each indicator across all countries, to judge the robustness of each indicator i.e. to indicate that some indicators are more difficult to rate than others;
- each country report across all indicators, to judge quality of the country assessment i.e. to indicate the varying skill of assessors at finding and presenting relevant evidence or the adequacy of resources allocated to the assessment or the impact of quality assurance arrangements.

4. All the reports received have been reviewed in detail by the Secretariat according to a standard process and recorded in a standard format. Most reports were reviewed by two assessors independently and the reviews then compared and reconciled. To ensure consistency only two individuals were involved for the Secretariat. A few reports were at the end of the process reviewed by only one assessor as some country reports were received close to the report completion deadline.

Assessor Variance

5. Assessor variance is the difference between the scores assigned to indicators in country reports, and the scores assigned, on the evidence available in the reports, by Secretariat’s assessors.

6. Variance is disaggregated into two types:

i. Evidence variance: Where the Secretariat does not dispute the score but the country report does not clearly disclose the evidence for the score. For instance, some country reports have merely highlighted or repeated the relevant requirements description from the Framework, with no evidence that the requirement is met. In other cases, the indicator has been rated without rating the underlying dimensions and with no clear presentation of evidence related to each dimension. Insufficient evidence on a dimension impacts this statistic only if it could change the indicator rating.

ii. Scoring variance: Where the Secretariat believes that a different score should have been assigned, either higher or lower, on the basis of the evidence presented in the country report.

7. As a variation of Scoring Variance, Weighted Scoring Variance is also measured. As changes of score may be large or small, they can be weighted as follows:

Single scale difference, eg. C vs. C+, or C vs. D+	1 point
Double scale difference, eg. C vs B, or C vs D	2 points
Triple scale difference, eg. C+ vs D, or C+ vs A	3 points
Etc.	

8. The index of 'weighted scoring variance' is the sum of points from the above divided by the number of indicators scored. This can be calculated by country and by indicator, as before.

9. The relationship between these measures is that

compliance index = 100% - Evidence variance – Scoring variance.

10. The indices explained above are shown for each indicator in the table below on the basis of 18 country PFM performance assessments of central government (i.e. excluding Bangladesh indicator ratings).

Analysis of Scoring Compliance, Assessor Variance and Coverage by Indicator

	Indicator	No. of cases correctly scored on sufficient evidence	Compliance index	Evidence variance	Scoring variance	Weighted scoring variance	Frequency of scoring indicator where applicable
A. PFM OUTTURNS Credibility of the budget							
PI-1	Aggregate expenditure out-turn compared to original approved budget	**9**	**53%**	**18%**	**29%**	**0.72**	**95%**
PI-2	Composition of expenditure out-turn compared to original approved budget	**6**	**35%**	**41%**	**24%**	**0.65**	**95%**
PI-3	Aggregate revenue out-turn compared to original approved budget	**12**	**75%**	**19%**	**6%**	**0.38**	**89%**
PI-4	Stock and monitoring of expenditure payment arrears	**7**	**44%**	**31%**	**25%**	**0.44**	**95%**
B. KEY CROSS-CUTTING ISSUES: Comprehensiveness and Transparency							
PI-5	Classification of the budget	**11**	**61%**	**28%**	**11%**	**0.17**	**100%**
PI-6	Comprehensiveness of information included in budget documentation	**14**	**78%**	**17%**	**6%**	**0.06**	**100%**
PI-7	Extent of unreported government operations	**1**	**6%**	**94%**	**0%**	**0.00**	**89%**
PI-8	Transparency of inter-governmental fiscal relations	**6**	**40%**	**53%**	**7%**	**0.07**	**89%**
PI-9	Oversight of aggregate fiscal risk from other public sector entities.	**10**	**59%**	**35%**	**6%**	**0.11**	**95%**
PI-10	Public access to key fiscal information	**12**	**67%**	**22%**	**11%**	**0.17**	**100%**
C. BUDGET CYCLE							
PI-11	Orderliness and participation in the annual budget process	**12**	**71%**	**6%**	**24%**	**0.29**	**95%**
PI-12	Multi-year perspective in fiscal planning, expenditure policy and budgeting	**10**	**63%**	**25%**	**13%**	**0.25**	**89%**
PI-13	Transparency of taxpayer obligations and liabilities	**7**	**47%**	**47%**	**7%**	**0.07**	**84%**
PI-14	Effectiveness of measures for taxpayer registration and tax assessment	**5**	**33%**	**47%**	**20%**	**0.14**	**79%**

	Indicator	No. of cases correctly scored on sufficient evidence	Compliance index	Evidence variance	Scoring variance	Weighted scoring variance	Frequency of scoring indicator where applicable
PI-15	Effectiveness in collection of tax payments	**6**	**43%**	**29%**	**29%**	**0.29**	**79%**
PI-16	Predictability in the availability of funds for commitment of expenditures	**10**	**56%**	**17%**	**28%**	**0.44**	**100%**
PI-17	Recording and management of cash balances, debt and guarantees	**6**	**38%**	**19%**	**44%**	**0.44**	**89%**
PI-18	Effectiveness of payroll controls	**8**	**53%**	**33%**	**13%**	**0.20**	**84%**
PI-19	Competition, value for money and controls in procurement	**9**	**56%**	**19%**	**25%**	**0.31**	**89%**
PI-20	Effectiveness of internal controls for non-salary expenditure	**7**	**41%**	**47%**	**12%**	**0.24**	**95%**
PI-21	Effectiveness of internal audit	**7**	**41%**	**47%**	**12%**	**0.18**	**95%**
C(iii) Accounting, Recording and Reporting							
PI-22	Timeliness and regularity of accounts reconciliation	**7**	**47%**	**40%**	**13%**	**0.13**	**84%**
PI-23	Availability of information on resources received by service delivery units	**9**	**60%**	**27%**	**13%**	**0.27**	**79%**
PI-24	Quality and timeliness of in-year budget reports	**8**	**44%**	**17%**	**39%**	**0.61**	**100%**
PI-25	Quality and timeliness of annual financial statements	**6**	**33%**	**28%**	**39%**	**0.50**	**100%**
C(iv) External Scrutiny and Audit							
PI-26	Scope, nature and follow-up of external audit	**6**	**35%**	**29%**	**35%**	**0.47**	**95%**
PI-27	Legislative scrutiny of the annual budget law	**11**	**65%**	**29%**	**6%**	**0.06**	**100%**
PI-28	Legislative scrutiny of external audit reports	**7**	**50%**	**36%**	**14%**	**0.14**	**94%**

	Indicator	No. of cases correctly scored on sufficient evidence	Compliance index	Evidence variance	Scoring variance	Weighted scoring variance	Frequency of scoring indicator where applicable
D. DONOR PRACTICES							
D-1	Predictability of Direct Budget Support	**2**	**20%**	**80%**	**0%**	**0.00**	**75%**
D-2	Financial information provided by donors for budgeting and reporting on project and program aid	**3**	**23%**	**46%**	**31%**	**0.31**	**74%**
D-3	Proportion of aid that is managed by use of national procedures	**7**	**64%**	**27%**	**9%**	**0.18**	**63%**
ALL	**Average of 31 indicators**	**7.8**	**48%**	**34%**	**18%**	**0.27**	**90%**

Annex 3

Distribution of Indicator Scores

as per Country Assessment Reports

Indicator		A	B+	B	C+	C	D+	D
A. PFM-OUT-TURNS: Credibility of the budget								
PI-1	Aggregate expenditure out-turn compared to original approved budget	50%	0%	11%	6%	28%	0%	6%
PI-2	Composition of expenditure out-turn compared to original approved budget	6%	0%	22%	0%	39%	6%	28%
PI-3	Aggregate revenue out-turn compared to original approved budget	72%	0%	11%	0%	11%	0%	6%
PI-4	Stock and monitoring of expenditure payment arrears	6%	0%	12%	0%	18%	24%	41%
B. KEY CROSS-CUTTING ISSUES: Comprehensiveness and Transparency								
PI-5	Classification of the budget	6%	0%	35%	6%	53%	0%	0%
PI-6	Comprehensiveness of information included in budget documentation	16%	0%	47%	5%	26%	0%	5%
PI-7	Extent of unreported government operations	12%	6%	18%	18%	18%	24%	6%
PI-8	Transparency of inter-governmental fiscal relations	6%	6%	0%	19%	19%	31%	19%
PI-9	Oversight of aggregate fiscal risk from other public sector entities.	0%	0%	0%	6%	39%	28%	28%
PI-10	Public access to key fiscal information	16%	0%	32%	0%	37%	5%	11%
C. BUDGET CYCLE								
PI-11	Orderliness and participation in the annual budget process	11%	6%	56%	6%	6%	11%	6%
PI-12	Multi-year perspective in fiscal planning, expenditure policy and budgeting	0%	0%	18%	12%	12%	41%	18%
PI-13	Transparency of taxpayer obligations and liabilities	6%	6%	31%	25%	13%	19%	0%
PI-14	Effectiveness of measures for taxpayer registration and tax assessment	7%	0%	0%	21%	29%	29%	14%
PI-15	Effectiveness in collection of tax payments	13%	13%	0%	0%	7%	47%	20%
PI-16	Predictability in the availability of funds for commitment of expenditures	5%	0%	11%	21%	5%	37%	21%

Indicator		A	B+	B	C+	C	D+	D
PI-17	Recording and management of cash balances, debt and guarantees	6%	24%	18%	24%	29%	0%	0%
PI-18	Effectiveness of payroll controls	0%	0%	6%	13%	13%	63%	6%
PI-19	Competition, value for money and controls in procurement	0%	7%	27%	13%	27%	27%	0%
PI-20	Effectiveness of internal controls for non-salary expenditure	0%	6%	0%	22%	22%	39%	11%
PI-21	Effectiveness of internal audit	0%	0%	0%	17%	17%	28%	39%
C(iii) Accounting, Recording and Reporting								
PI-22	Timeliness and regularity of accounts reconciliation	0%	0%	19%	44%	13%	6%	19%
PI-23	Availability of information on resources received by service delivery units	13%	0%	13%	0%	33%	0%	40%
PI-24	Quality and timeliness of in-year budget reports	5%	5%	5%	37%	16%	16%	16%
PI-25	Quality and timeliness of annual financial statements	0%	5%	5%	11%	21%	32%	26%
PI-26	Scope, nature and follow-up of external audit	0%	6%	0%	28%	11%	22%	33%
PI-27	Legislative scrutiny of the annual budget law	0%	16%	0%	37%	11%	26%	11%
PI-28	Legislative scrutiny of external audit reports	0%	0%	6%	13%	25%	31%	25%
D. DONOR PRACTICES								
D-1	Predictability of Direct Budget Support	25%	0%	8%	17%	17%	17%	17%
D-2	Financial information provided by donors for budgeting and reporting on project and program aid	14%	0%	0%	7%	21%	36%	21%
D-3	Proportion of aid that is managed by use of national procedures	0%	0%	8%	0%	42%	0%	50%
ALL	**Average of 31 indicators**	**10%**	**3%**	**14%**	**14%**	**22%**	**21%**	**17%**

Annex 4

Applications of the PEFA PFM Performance Measurement Framework at Sub-National Government Level

	Ethiopia	**Mexico states**	**Pakistan Punjab State**	**Uganda**
Coverage of SN government	7 SN governments assessed (98% of national population) and integrated with federal government assessment.	8 state governments on voluntary basis (may gradually expand)	1 state government assessed (population 80 mill)	6 SN governments assessed (10% of national population) in parallel with central government assessment.
Report	Ethiopia - Fiduciary Assessment (one report covering both federal and state government assessments)	State-Level Technical Assistance to Establish Performance Indicators for Procurement, Public Expenditure and Financial Management	Punjab Province Pakistan; Public Financial Management and Accountability Assessment,	Local Government Public Financial Management Assessment 2005 + One separate report for each of six LGs assessed
Date	Final Draft, Feb 6, 2005	Progress Report June 2005; work continuing 2006	Draft 16 February 2005 for workshop review.	Draft, Dec 21, 2005
Lead Agency	DFID / EC	World Bank	World Bank	EC
Other donors involved	-	IADB	EC, DFID	Financing by EC, Sweden, Norway, UK, Ireland
Government involvement	Data provision and workshop on provisional findings ??	Self-assessment by federal states on voluntary basis	Provided general data inputs, partial self-assessments, and participated in workshops at planning and draft rep. stages.	

	Ethiopia	Mexico states	Pakistan Punjab State	Uganda
Purpose	Fiduciary risk assessment for the Consortium of Budget Support Donors in Ethiopia.	Provide state governments with a tool for benchmarking management performance and promote continuous enhancement of institutional capabilities	Create a foundation for the government's commitment to a program of institutional reorganization and legal reform	To feedback on fiduciary risk at and PFM reform impact at LG level. Feed into combined Central and Local Govt Assessment and Review of Progress on CIFA Action Plan
Use of PEFA Framework	Used 14 of the PEFA indicators (v.2004)	18 PEFA indicators (mixed versions 2004 and 2005) have been used to various degree as part of the 42 indicators in the proposed set	The 28 government performance indicators used (v. 2004) with a few special interpretations to reflect the sub-national status of the assessed entity.	All 28+3 standard indicators (version June 2005) used with minimal changes to fit LG situation
Main deviations from PEFA Framework	14+2 indicators not used. Scoring on 3-point ordinal scale (DFID FRA system), not using PEFA calibration.	The proposed set goes far beyond the PEFA scope both in breath (e.g. allocative efficiency) and depth (e.g. 17 procurement indicators). No scoring applied, but possibly one benchmark level. Only factual data to be presented. Qualitative dimensions avoided.	No use of donor indicators.	5 additional indicators for CG-LG relationship

Annex 5.

Checklist for terms of reference for undertaking a PEFA assessment and preparing a PFM Performance Report

This document seeks to outline the areas that terms of reference (TOR) for undertaking a PEFA assessment and preparing a PFM Performance Report would need to address, whilst noting that it will be necessary to tailor the content and structure of the terms of reference to the country circumstances. It is expected that the TOR would be discussed by the donor group and the government.

1. Background and context

Cover issues such as:

- broader development context;
- history of engagement on PFM agenda;
- status of government's PFM reform agenda;
- relevant prior diagnostics;
- the current process of engagement, analysis, support and monitoring that is taking place on PFM between government and donors and the link to relevant donor operations (such as budget support); and
- the role of the planned assessment, how it fits within the overall engagement and how it contributes to harmonized and rationalized PFM work.

2. Purpose of the assessment

A concise statement of the purpose of the assessment, within the context described above, and how it is to be used in the dialogue regarding PFM.

3. Involvement of stakeholders in the assessment

- Identification of how relevant stakeholders are to be involved in the assessment, reflecting the requirement that it must be coordinated amongst donors. This would include identification of lead donor(s) and team members
- Involvement of government in the assessment including identification of any government liaison official or government team members
- Involvement of wider donor group
- Identification of any other stakeholders that may be involved (e.g. the Supreme Audit Institution)
- Use and funding of consultants in undertaking the assessment
- Organizational arrangements for participation and involvement, including a possible pre-mission workshop for government and donor officials for briefing, team-building and detailed planning of the work; and a possible restitution workshop for discussion of findings and their implications for PFM reforms.

4. Methodology for undertaking the assessment

- Description of the coordination of the assessment with any relevant, related PFM work, and to relevant donor operations.
- Reference to the use of the PEFA PFM Performance Measurement Framework and the requirement to apply the Framework guidance contained in the Framework's annexes 1 and 2. Reference to any additions that have been agreed.
- Any need to track progress from a previous PEFA based assessment or from an assessment of HIPC Expenditure Tracking indicators (which could imply additional work for the team in order to explain in detail the changes in performance and the implications for indicator ratings).
- Sources of information for the scoring of the indicators (including relevant prior diagnostics) and means for collecting information and evidence. Reference to the need for the assessment team to highlight information gaps, rather than to attempt to give a score where data is substantively incomplete.
- Arrangements for meetings and interviews, including the government's role in data collection before and during the field mission(s).
- Arrangements for external validation/quality assurance.

5. Reporting

Identification of to whom the draft report should be submitted, to whom it should be circulated for comment, and how comments are to be processed.

6. Consultation and follow up to the assessment

Identification of how the report will be discussed and used in the engagement on the PFM agenda, any follow up that may be expected to the completion of the assessment, and the expectation regarding the frequency in undertaking the assessment in future.

7. Specific steps, timetable and deliverables

Drawing from the previous sections, identification of the specific steps to be undertaken, outputs to be produced in connection with each stage of the study and the timetable for these. This should include number and duration of field missions and the timing and purpose of any workshops prior to, during and after the field work.

8. Team composition and Inputs

- Identification of the staff and consultants' inputs required in the team conducting the assessment, covering the skills required to cover the range of PFM issues and how local knowledge will be utilized (to be linked closely to the features of the TOR above as well as country specific features such as the ease with which data may be obtained and the field mission logistics).
- Specification of expected period of the field mission(s) and staff time involved as well as time set aside for home office inputs before and/or after field missions.
- Internal team management arrangements, particularly where a team is composed of members who are not under a unified contractual arrangement.

Annex 6

Abbreviations and Acronyms

AAP	Assessment and Action Plan
AGA	Autonomous Government Agencies
AsDB	Asian Development Bank
CAS	Country Assistance Strategy
CFAA	Country Financial Accountability Assessment
CIFA	Country Integrated Fiduciary Assessment
COFOG	Classifications of Functions of Government
CPAR	Country Procurement Assessment Report
DAC	Development Assistance Committee of OECD
DFID	United Kingdom Department for International Development
EC	European Commission
ECA	Eastern Europe and Central Asia
FINMAP	Uganda Financial Sector Project
FM	Financial Management
FMS	Financial Management Specialists
FRA	Fiduciary Risk Assessment
GBS	General Budget Support
HIPC	Highly Indebted Poor Countries
IADB	Inter-American Development Bank
IFAC	International Federation of Accountants
IMF	International Monetary Fund
GFS	Government Finance Statistics
INTOSAI	International Organization of Supreme Audit Institutions
LAC	Latin America and the Caribbean
LG	Local Government
IPSAS	International Public Sector Accounting Standards (of IFAC)
LGA	Local Government Authority
MDA	Ministries, Departments and Agencies
MENA	Middle East and North Africa

MOF	Ministry of Finance
NGO	Non Governmental Organization
PEFA	Public Expenditure and Financial Accountability
PEFAR	Public Expenditure and Financial Accountability Review
PER	Public Expenditure Review
PFM	Public Financial Management
PFM-PR	PFM Performance Report
PI	Performance Indicator
PMF	Performance Measurement Framework
PNG	Papua New Guinea
PPER	Project Performance Evaluation Report
Q&A	Question and Answer
QAG	Quality Assurance Group
SAI	Supreme Audit Institution
SECO	Swiss State Secretariat for Economic Affairs
SN	Sub-National (government)
TOR	Terms of Reference
UN	United Nations
WB	World Bank