



Good Practice when Undertaking a Repeat Assessment

Guidance for Assessment Planners and Assessors

Final

PEFA Secretariat
February 1, 2010

Executive Summary

The Public Expenditure and Financial Accountability (PEFA) Performance Measurement Framework has been used for more than 150 assessments of a country's PFM system since its introduction in 2005. Although many recent assessments are 'repeats', they frequently fail to satisfy one of the key intentions of the Framework: to provide a clear picture of specific changes in performance since the initial assessment.

Hence the Secretariat has been requested to produce guidance on 'good practice' and this note makes suggestions for each stage of a typical repeat assessment.

In summary, the guidance is as follows:

FOR ASSESSMENT MANAGERS/SPONSORS

CONCEPT NOTE/TERMS OF REFERENCE

- Don't undertake the assessment too frequently: change takes time, and the recommended interval for repeat use of the PEFA Framework **is at least three years**.
- Ensure that the **CN/TORs are sufficiently detailed**, and are clearly understood by all Stakeholders.
- **Facilitate access to documentation** from the previous assessment, whether published or not, and to unpublished notes and if necessary, verbal inputs from the leader of the previous assessment team.
- The **lead agency of the earlier assessment** (if different from that leading the repeat) should be part of the reference group for the repeat exercise.
- It would be **an advantage to use the same team** as previously, provided that team produced a quality report and that the team members have not in the meantime been engaged in designing and implementing PFM reform and capacity development in the country, as this could present a conflict of interest in the assessment.

FOR ASSESSORS

PREPARATION

- **Contact the team leader from the earlier assessment** to obtain additional information on data and sources as required (including comments from peer reviewers).
- Don't approach the assessment with the **assumption that your predecessors were either wrong in their ratings, nor that they were perfectly correct!**
- **Use the previous assessment as the starting point** when explaining the trajectory of change in the performance of the PFM system (rather than examining each indicator in isolation from the previous assessment and then comparing the ratings, as this runs the risk of failing to identify differences in definitions, etc.).

FIELDWORK

- **Verify the basis on which the earlier score was assigned** and identify any indicators for which inadequate information is available for this verification.
- **Keep comprehensive notes on data and sources** for each rating – for future tracking (even if this has to be in a confidential file that will not be included in the report proper).
- Do not use “D” scoring when insufficient information is available to assign a score.
- ***Do not attempt to re-rate the earlier assessment:*** these scores have been incorporated in documents, which – whether published or not – various stakeholders will have used.

DRAFTING THE REPORT

- Should mistakes be found in the earlier ratings, explain them in the discussion of changes in indicator scores. **A note to the indicator overview table may be inserted to explain that the present and previous ratings are not comparable.**
- **Be precise about data sources** (to aid successors).
- **Include the CN/ToR as an Annex** to the report.
- **In the Summary Assessment:**
 - **Present an overview of changes in performance** ratings since the earlier report and a table with scores for N then N+1;
 - **List how many indicators** have deteriorated, how many have improved and how many remain unchanged;
 - **Do not aggregate results from n to n+1 to show a % improvement!**
- **In the narrative:**
 - Provide **an overview of problems with old scores** from data or definitional changes;
 - **Identify the impact of disagreements** with scores from the original assessment;
 - **Identify the reasons for changes in scores between the two assessments, indicator by indicator**, so that the reader can understand in detail what has happened.

1 Purpose of this note

Since 2005, the Public Expenditure and Financial Accountability (PEFA) Performance Measurement Framework has been used to conduct assessments of the Public Financial Management systems in more than one hundred countries. So far, most of the assessments undertaken have been baseline assessments, but as might be expected, repeat assessments are becoming more common as in some regions, baseline coverage is almost complete and also, nearly three years have passed since the first assessment took place.

Although an increasing proportion of assessments are repeats, they often fail to satisfy one of the key intentions of the Framework: that of providing a clear picture of specific changes in system performance since the initial assessment. In some cases, progress is described in the Summary Assessment but is not tracked in detail at the level of individual indicators.

As more and more repeat assessments are being conducted, Assessors, Government Officials and representatives of Development Agencies have requested the Secretariat to produce guidance on ‘good practice’ when undertaking a repeat assessment, and hence this note has been prepared.

2 Context: use of the Framework

The PFM Performance Measurement Framework is one of the three elements of the ‘Strengthened Approach’ to supporting PFM reforms: these are: (i) a country-led PFM reform strategy and action plan, (ii) a coordinated donor multi-year program of PFM work that supports and is aligned with the government’s PFM reform strategy and, (iii) a common information pool.

The Framework is a tool for achieving the third of the elements – a common information pool – and in particular has been designed to measure the PFM performance of countries at various levels of development over time, by allowing the establishment of a baseline against which progress can be assessed.

When used consistently, stakeholders can expect that the repeated application of the Framework will provide evidence of the extent to which country PFM performance is improving, or not. In addition, the PFM-Performance Report will recognize the efforts made by a government to reform its PFM system by describing recent and on-going reform measures, although these may not yet have impacted on PFM performance.

3 Consistency

It is immediately apparent that for two assessments conducted some time apart to be comparable, they must represent consistent applications of the methodology. Assuming that the initial assessment represents a rigorous application of the methodology,

stakeholders would ideally want to know that progress has been tracked accurately over time by using the:-

- SAME Assessors;
- SAME Sponsor;
- SAME Counterparts;
- SAME Data sources.

However, it is unlikely that all, and perhaps even any of the above elements will be available for a repeat assessment. For example, if consultants were employed for the previous assessment, the same team is unlikely to be available three years later (by the very nature of consultants' work patterns). Similarly, some government officials will certainly have moved positions over a period of years. In addition, less than half of the repeat exercises have been sponsored by the same development agency, and some data sources, such as comments by Sponsors and Peer Reviewers (including the Secretariat) on draft versions of the previous assessment report may fall foul of confidentiality issues.

There is also the possibility that the previous assessment has not been perceived as being of good quality, in which case a different question arises: does this assessment provide the reliable baseline so necessary if progress is to be meaningfully tracked, and if not, how should the repeat assessors proceed?

In the light of these practical issues, the question of what is "good practice" for assessors when undertaking a repeat assessment becomes crucial.

4 Issues identified in Repeat Assessments

The main purpose of a repeat assessment is to track performance since the previous assessment. In order to do so, it is necessary to make sure that like is compared with like. This means understanding the evidence upon which the scores in the previous assessment were rated. This may not always be obvious: for example, revenue and expenditure figures available to the earlier assessors may have changed when audited, and hence the repeat assessors may be given different fiscal data for the same year. These and other matters are considered in more detail below.

In considering the twenty-plus repeat assessments completed to date, two points are immediately apparent. Firstly, nearly all these assessment were completed in less than the time period suggested in the PEFA Framework (3-5 years) even though it is widely understood that it normally takes time for reform efforts to show impact. Secondly, the quality of these repeat assessments varies. While some repeat assessments are of good quality, others are not. There are several factors that might explain these deficiencies:

- Firstly, despite the assessment team being aware that they were conducting a repeat exercise, they may have taken a deliberate decision not to access the previous report, as witnessed in this quotation taken from an assessment report: *"The team wanted to*

reach independent conclusions without being influenced by previous assessments. This was discussed with stakeholders during the assessment.” In this situation, explaining the reasons for changes in ratings becomes much more difficult.

- Secondly, the Concept Note/Terms of Reference may not have been sufficiently detailed in specifying what the assessors were expected to do in tracking progress.
- Thirdly, assessors may not have seen the previous assessment, and if this was sponsored by a different lead Agency, may not have had access to draft documents or even the final report if this was not made public: hence they feel unable to attempt a comparison.
- Fourthly, assessment teams may not have access to comments of Peer Reviewers (including the Secretariat) on the *draft* report of the previous assessment, which might have identified scoring issues that were then **not** taken into account in the *final* version of the report, and hence may not have all the relevant information necessary to attempt a comparison.
- Fifthly, assessment teams may have a different view of the correctness of the score from the previous assessment, and hence be unable to make a meaningful comparison.
- Finally, some repeat assessments represent an attempt to create a more accurate baseline than that achieved by the previous assessment, in which case a comparison might be deemed meaningless.

5 Examples of Good practice

The Secretariat would define a satisfactory repeat assessment as one that verifies the basis on which the earlier score was assigned and identifies any indicators for which inadequate information was available for this verification. In addition, a satisfactory repeat assessment would have to consider any obvious mistakes in the use of evidence for the previous score in order to be able to compare like with like.

Examples of common weaknesses identified in the Secretariat’s comments taken from reviews of recent draft reports would be:

- *“There is a lack of emphasis accorded to the tracking of progress. The report did not outline the factors explaining progress under each indicator in addition to providing the comparative scores.”*
- *“The lack of a comparison of the 200X and 200Y ratings by indicator (and possibly sub-indicator) does not provide an understanding of changes (and lack of changes).”*
- *“The comparison or explanation is incorporated for only a few indicators.”*
- *“There is no overview table showing previous & current scores, together with explanations for progress or lack of it.”*

The sort of issues listed above would be addressed if:

- the summary assessment provides a brief overview of changes in performance ratings since the earlier report, and a table with both sets of ratings;
- the reasons for changes in scores between the two assessments are outlined, indicator by indicator;
- a detailed overview of progress between the two assessments is included as an Annex.

Indicator scores will provide a crude overview of changes over time, but individual dimensions may change differently, and performance may not change enough to change the indicator score (in which case it may be appropriate to use an arrow). Hence a more detailed explanation will be required.

To illustrate good practice, the Secretariat offers the following examples taken from recent reports:

- In the Summary Assessment:

..... "In Table A1, 7 indicators appear to have deteriorated since 2005. In fact, some of the ratings in 2005 now appear overrated, as explained in the comparison tables throughout the report. Only one indicator, D-1, appears to have really deteriorated. This is surprising, given the templates developed by the Reform Program, and needs to be followed up in the context of the review. Twelve indicators appear to have remained unchanged, though in some (PI-10, 15, 18, 26, and 27) there have been significant reforms, though not sufficient to change their ratings. Another 12 indicators appear to have improved, though a re-rating of PI-5 in 200X reduces this to 11 (PI-4, 6, 9, 13, 14, 17, 20, 21, 22, 24, and D-2). Some of these reflect the rollout and increasing use of the new FMS and other reforms. Several weaknesses, however, remain to be addressed more effectively, as shown in the summary."

Table A1			
		Score 200X	Score 200Y
PI-1	<i>Aggregate expenditure out-turn compared to original approved budget</i>	B	B
PI-2	<i>Composition of expenditure out-turn compared to original approved budget</i>	C	C
PI-3	<i>Aggregate revenue out-turn compared to original approved budget</i>	A	A
PI-4	<i>Stock and monitoring of expenditure payment arrears</i>	D	D+

- For individual indicators:

“Comparison of 2004 & 2006 – Based on the information available in 2007, we note that the score established for indicator PI-1 in 2004 was in fact erroneous, in particular due to the inclusion within the calculation of expenditure on debt servicing. At that time, the expenditure deviations registered were +3.7% in 2002, -10.8 % in 2003 (due to the floods which ravaged the country in 2000 and 2001, necessitating heavy relief and reconstruction expenditures which were difficult to budget accurately) and -3.7% in 2004. Our updated analysis of the data for the year 2004 (based on the final statement of accounts) shows a deviation in the 2004 fiscal year of -5.19%, which would have given a “B” score.”

- Then, an Annex with a detailed overview providing the score PEFA N and PEFA N+1 for each indicator and a description of progress between N and N+1:

		Score 200X	Score 200Y	Changes between 0X & 0Y
PI-1	Aggregate expenditure out-turn compared to original approved budget	C	B	A continued focus on maintaining fiscal discipline has resulted in improvements in the predictability of the overall budget
PI-2	Composition of expenditure out-turn compared to original approved budget	D	D	There has been a modest improvement in the average variance in MDAs outturns above total expenditure deviations. This may reflect both sustained financing of priority policy programs as well as improvements in the release of resources to MDAs throughout the year. Nevertheless, budget variance remains high.
PI-3	Aggregate revenue out-turn compared to original approved budget	A	A	Continued conservative revenue projections and enhanced enforcement measures have kept revenue collections high relative to budgeted amounts.
PI-4	Stock and monitoring of expenditure payment arrears	D+	B+	
(i)	<i>Stock of expenditure payment arrears & recent change in stock</i>	<i>D</i>	<i>B</i>	Governments 5-year strategy for reducing expenditure payment arrears has resulted in year-on-year reductions in overall arrears

(ii)	Availability of data for monitoring the stock of expenditure payment arrears	B	A	Regular quarterly reports are now compiled by the Controller of Internal Audit
------	--	---	---	--

6 Disagreement with previous ratings

In order to be able to properly track and understand PFM performance changes and to have confidence in the comparison of ratings of an indicator over time, it is necessary for readers to understand from the narrative of the PFM-Performance Report what has changed, and by how much.

It must also be remembered that not all changes in scores necessarily represent a change in PFM system performance. Possible reasons for changes (taken from recent assessments) are as follows:

- **Different access to information**

Example from PI-7: *“This time a study of quasi-fiscal operations by parastatal enterprises has been taken into consideration. Such a study was not available for the earlier assessment. Hence a lower score this time, but no evidence that anything has deteriorated”.*

Another example: PI-4: *“This time the assessment is based on a different definition of arrears, last time the assessment was based on the accounting definition and hence only included payments which had been accepted and processed for payment by the Treasury. So the volume of arrears has increased, but performance may not have deteriorated.”*

- **Different sampling**

Example PI-21: *“In an earlier report the assessment was based on information on the state of affairs in the Ministries of Agriculture, Justice and Finance. This time it is based on information from Ministries of Education and Justice. It is therefore not certain if a direct comparison is valid.”*

- **Different interpretation**

Example PI-24(ii): *“During the last year issue of budget execution reports took place 4-7 weeks after end of quarter, with average 5.5 weeks: given a ‘B’ rating. At the last assessment the performance was exactly the same but given ‘C’ because two reports were more than 6 weeks delayed.”*

- **Scoring methodology – mistakes do happen**

The influence of mistakes on changes in scores must be explained. Again, a few examples:

- *“wrong assignment of score despite very clear evidence”;*

- “use of M1 scoring aggregation of dimensions where M2 should have been used”;
- “assignment of a ‘+’ to a single dimension indicator”.

The Secretariat **recommends** that should repeat assessors find any of the above issues, they should **not attempt an explicit re-rating of the earlier assessment**, as this work may have been utilized and endless debates may follow. If indisputable mistakes in rating are found in earlier work, they can be explained as part of the discussion of changes in indicator ratings.

A note to the indicator overview table may be inserted to explain that the present and previous ratings are not comparable, or how a different view of the performance at the earlier assessment may have influenced the conclusions about the direction of change in PFM system performance since then.

It is also important to note that the assessment of each indicator should be linked to the previous assessment i.e. that the focus is on explaining the performance trajectory by using the earlier assessment as a starting point, rather than doing an assessment of the indicator isolated from the previous assessment then subsequently comparing the ratings: the latter approach risks ignoring the sort of differences highlighted above.

Finally, it may be desirable but not always possible to say why performance improved – as this may require more detailed analysis than the PFM-PR is meant to provide, by digging into the underlying causes (such as capacity factors).

7 Reporting on individual indicators

The Secretariat recommends that this may usefully be done in a table format, as shown below:

Indicator	Score 2006	Score 2010	Performance change	Other factors
PI-1	C	B	Performance appears improved based on 2006: 6%, 11%, 18% 2010: 5%, 11%, 6%	Not clear if all external project funds were excluded from data for 2006 assessment but probably insignificant issue.
PI-4 (i)	A	C	Performance change is uncertain, despite reported arrears increase from 1% in 2006 to 6% in 2010.	2006 assessment used data on pending payment orders only, not overdue invoices.

8 Conclusions on overall changes in PFM system performance

As the number of repeat assessments has increased, there have been several attempts by authors to produce an aggregated measure of performance changes. Such an aggregated measure may be particularly tempting in cases where some indicators show progress whereas others indicate deterioration in performance. Aggregation typically involves both the conversion of ordinal indicator ratings to numerical values and the allocation of weights to the individual indicators. However, as there is no scientifically correct method to undertake these conversions, the Secretariat neither supports nor recommends any particular approach. (For a detailed discussion of the methodological issues that arise when attempting to aggregate PEFA scores, a note “*Issues in Comparison and Aggregation of PEFA Assessment Results Over Time and Across Countries*” is available on the PEFA website (www.pefa.org).

As is clear from this Guidance Note, the recommended approach is to compare scores for each indicator and use the narrative in the PFM-Performance Report to draw conclusions on reasons for differences in scores and hence any change in performance of the PFM system. The Summary Assessment section should be used to draw overall conclusions on the direction and extent of change, highlighting the relative importance of various indicator changes and links between them, as appropriate.

9 The Concept Note/Terms of Reference

Tracking progress over time should be treated carefully and precisely in the Concept Note/Terms of Reference (CN/ToR). In order to be able to track properly and understand PFM performance changes and to have confidence in the comparison of ratings of an indicator over time, it must be possible to understand from the narrative of the report what has changed and by how much.

Hence assessors must be aware that the comparison of assessments and tracking of performance over time (with explanations) is expected of them. The CN/ToR should require that the team explain all factors that impact a change in rating, indicator by indicator; identify the performance change and ensure that any reader can track the change from the previous assessment. This can be a substantial part of the report writing, even if the existence of a previous (quality) assessment would normally be an advantage in the assessment work.

For a useful trajectory of progress to emerge from the repeat exercise it is important that the assessment of each indicator is linked to the previous assessment i.e. that the focus is on explaining the performance trajectory by using the earlier assessment as a starting point, rather than doing an assessment of the indicator isolated from the previous assessment and then comparing the ratings afterwards (the danger of this approach is that the trajectory may become subject to differences in sampling of definitions, new information becoming available or different interpretation or judgment in borderline

cases). The CN/ToR should be explicit on this point, as illustrated in the example at **ANNEX A**.

An example of a Concept Note/Terms of Reference that is **likely** to produce satisfactory repeat assessment:

- *“The objective of the assessment mission is to draft a comprehensive “PFM – Performance Report” (PFM-PR) prepared according to the PEFA methodology, so as to provide an analysis of the overall performance of the PFM systems of the country as well as to follow-up on progress against the PEFA indicators from the previous assessment that will permit the measuring over time of changes in performance.*
- *The specific objectives of this assignment are to:*
 1. *Update the overview of PFM performance in accordance with the PEFA Performance Measurement Framework.*
 2. *Establish and explain the level of improvement in performance based on the PEFA indicators scores by comparison to the results found during the previous evaluation.*
 3. *Assess the results of the PEFA review i.e. the performance change in relation to the project activities and possible effects on the scores attained.*
- *The assessors will be provided with:*
 - a) *The final report of the previous assessment*
 - b) *All peer review comments on the previous assessment*
 - c) *Access to the Leader of the previous assessment Team to explain details of the ratings allocated should this be necessary”.*

An extract from a Concept Note/Terms of Reference that is **unlikely** to produce a satisfactory repeat assessment:

- *“There is a need to take stock of progress since the previous exercise” (but nothing more specific is expected from the assessors, such as even looking at previous scores).*

Example: Concept Note/Terms of Reference

Specific objectives

1. To update the overview of PFM performance in accordance with the PEFA Performance Measurement Framework.
2. Establish and explain the level of improvement in performance based on the PEFA indicators scores by comparison to the results found during the previous evaluation.
3. An assessment of the results of the PEFA review i.e. performance change in relation to the reform program component activities and possible effects on the scores attained.
4. Cognizance should be taken possible reasons that could have contributed to the change in scores such as the following:
 - Changes in definitions
 - Improved availability of or access to information
 - Different information sampling and aggregation
 - Different approach to professional judgments
 - Scoring methodology inaccuracies in previous assessment such as the use of “D” scoring when insufficient information was available to assign a score.
5. The consultants in the report should ensure that:
 - All factors that impact a change in rating indicator-by-indicator are explained
 - The performance change is identified
 - Any reader can track the change from the previous assessment.