

**Draft Policy Brief:**

**Revised Indicator 9a for the Global Partnership Monitoring  
Framework**

*March 2015*



This policy brief has been produced with the kind assistance of the European Union and the German Ministry for Economic Cooperation and Development (BMZ) through the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. GIZ does not necessarily share the views expressed in this publication. Responsibility for this content rests entirely with CABRI and the Effective Institutions Platform.

## Introduction

This Policy Brief sets out proposals arising out of the work of Effective Institutions Platform (EIP) to formulate a revised indicator 9a for the Global Partnership Monitoring Framework (GPMF). The Policy Brief and the process to develop it should be seen as a first step in a broader global discussion to revise indicator 9a.

## Background

Effective public financial management (PFM) systems are necessary to ensure the legality and propriety of the use of public money, as well as its effectiveness, efficiency and economy. It is essential to the achievement of development objectives. At the same time, an effective and transparent PFM systems provides donors with assurance on the use of their funds that are managed through country systems.

The importance of effective PFM systems for effective development cooperation was recognised in the Paris Declaration (paragraphs 17-30), and reconfirmed in the Busan Global Partnership agreement (paragraph 19). These agreements commit countries to strengthen their PFM systems, and donors to use the strengthened systems.

The Busan Partnership agreement is supported through the GPMF. The framework enables tracking progress against the commitments and actions of the agreement, including on the use of country systems (UCS). Indicator 9 (see Box 1) in the framework measures whether partner country systems are strengthened and used (respectively through indicator 9a: Quality of Developing country PFM systems and 9b: Use of Country Procurement and PFM systems). Ideally, improvements in 9a should be followed by improvements in 9b. This Policy Brief discusses adjustments to indicator 9a. None of the proposals concern the calculation of 9b.

### Box 1. Indicator 9 in brief

Indicator 9 is one of the indicators in the framework that is based on the Paris Declaration Monitoring Framework (PDMF). The current 9a is taken over from the PDMF, which used the World Bank Policy and Institutional Assessment (CPIA) results for one criterion, Indicator 13, a measure of the quality of country's budget and financial management systems. Specifically the following dimensions are rated by World Bank staff to compute Indicator 13:

- a comprehensive and credible budget, linked to policy priorities;
- effective financial management systems to ensure that the budget is implemented as intended in a controlled and predictable way; and
- timely and accurate accounting and fiscal reporting, including timely and audited public accounts and effective arrangements for follow up.

The GPMF 9b indicator was also taken over from the PDMF, but combining indicators 5a (use of country PFM systems) and 5b (use of country procurement systems) to offer a single composite indicator. The 2014 Global Partnership Progress Report<sup>1</sup> found no change overall in the quality or use of country's PFM systems using these indicators.

The need for an adjustment to indicator 9a arose out of concerns that the current indicator does not sufficiently reflect or satisfy many of the principles on which the Busan agreement is predicated,

<sup>1</sup> OECD/UNDP (2014). Making development cooperation more effective: 2014 Progress Report. OECD Publishing

namely: ownership of development priorities by developing countries, a focus on results, and transparency and mutual accountability.

### Key processes to develop a revised indicator

The EIP took on adjustment of indicator 9a as part of its pillar on indicators for success. CABRI and GiZ on behalf of BMZ, respectively co-chair and member of the EIP, assumed responsibility for the work. The proposals put forward in this Policy Brief are the outcome of this process.

The work to develop the revised proposal occurred in three phases: initial research and recommendations on the form and content of a revised indicator; consultation on the recommendations and the preparation of a proposal; and consultation on the proposal.

The base research was undertaken between January and April 2014, and resulted in a research paper and brief which was presented to EIP members at the High Level meeting of the Global Partnership in April 2014. The research comprised a literature review of existing PFM evaluation frameworks and indicators (see Box 2); a consideration of what is meant by quality PFM in the context of UCS; an elaboration of methodological options for a revised indicator and criteria for assessing the options; an assessment of options, including the use of existing PFM evaluation frameworks; a discussion on potential content for an indicator given the discussion on quality PFM; and recommendations on a revised indicator for further deliberation.

Following the consultation with EIP members on the research and recommendations, a proposal for a revised indicator was drafted for wider consultation. Besides the further discussion at the EIP Annual Meeting held in October 2014 in Paris, direct feedback was also received from 6 partner countries, 3 independent institutions and 5 donors.

### Box 2. Existing methodologies assessed

The research assessed the following widely used PFM diagnostic and evaluation frameworks for the suitability of their coverage and design for use in a revised PFM indicator.

- **Public Expenditure and Financial Accountability framework:** PEFA is managed by a partnership between development partners and includes the provision of reliable information on country PFM performance in its objectives. The revised PEFA framework comprises 30 indicators on country systems processes, outputs and outcomes (compared to 28 in the previous version). PEFA assessments are undertaken at the country level by an independent team, and quality assured by the PEFA Secretariat. Scoring is based on objective criteria, with quantitative elements where possible. There is no summary score. Repeat assessments occur every 3 to 5 years, and almost all developing countries have been assessed. PEFA assessments are only published if the country agrees.
- **CPIA:** CPIAs are undertaken by World Bank staff who score countries against 16 indicators on a six-point scale (1 low/weak, 6 high/strong). The scores are summarised to a single rating score, to enable comparisons and ranking across countries. The scoring process includes country consultation, and all scores and summary scores are published. The CPIA focuses on institutions (and less on outcomes as these are externally influenced), and is undertaken once a year. Please refer to box 1 for the dimensions rated in the CPIA and used in the current indicator.
- **Open Budget Index and Survey:** The OBI is focused on budget transparency and accountability and undertaken by country-based researchers, but quality controlled by the International Budget Partnership, a global independent NGO. The 2012 Survey comprised

125 questions across the budget cycle with specific scoring guidance, and submission of evidence. The Survey was undertaken in a 100 countries. The survey is repeated every second year.

- **DFID Fiduciary Risk Assessment:** DFID fiduciary risk assessment is tailored for quality PFM in the context of UCS. It is done by assessment teams and comprises 8 good practice principles (GPPs) and 15 benchmarks, which measures current status and trajectory of change. It uses PFM indicators within these where possible. It rates fiduciary and corruption risk as Low/Moderate /Substantial or High, as well as partner government commitment to improving PFM, strengthening domestic financial accountability and fighting corruption as credible, mixed commitment or not credible. Assessments are done every three years, and their results are not published.

The research found that while the existing methodologies include individual measures that can be used in a revised indicator 9a, and include desirable elements in terms of structure and measurement methodology, not one is fully satisfactory to use as Indicator 9a as is.

### Key assumptions and criteria

Key assumptions for the research and proposal were that the proper assessment of the quality of PFM systems should cover both the underlying PFM processes and outputs (for example a budget process and an executive budget proposal), and the PFM outcome of the system (for example acceptable budget variance). In other words, Indicator 9a should provide assurance that key PFM processes and outputs are in place, and that they result in a functioning PFM system. It was also assumed that “PFM systems” in a UCS context should cover systems which manage the core elements of the PFM cycle, namely budget preparation, budget execution, accounting and reporting, and accountability and audit, even if there are important questions to be addressed about the scope of coverage and emphasis on each system.

The high-level criteria for assessing the methodological options and indicator content were that: the indicator and its calculation had to be transparent; it had to provide a credible measure of a country’s PFM performance within the context of UCS; it had to provide consistency of measurement within countries over time to track progress; it had to be reliable and consistent in its measurement across countries; it had to be owned by countries and be sensitive to country context; it had to support mutual accountability; and it had to be cost-effective, preferably using existing measures so as not to increase transaction costs.

### Proposal for a revised Indicator 9a

The design of an indicator comprises three inter-dependent aspects: the structure of the indicator (how the indicator is structured, for example is it a single measure, a composite measure or a framework of measures); the scoring of the indicator (the measurement methodology used to make judgements on the measure(s) and whether a summary measure is calculated and how); and the content of the indicator (which aspects of PFM are being measured). Using the assumptions and criteria detailed above, this section sets out a proposal across these aspects. The discussion notes aspects of the proposal on which there was broad consensus in consultation, and aspects on which opinions from respondents diverged significantly.

### Construction of the indicator

It is proposed that the indicator comprises two components.

Component 1 (fixed component) will consist of a set of measures that is applied in all countries, and which reflects country performance on the core PFM systems that are deemed critical by

development partners to use country systems across countries. There will be 7 measures in this component.

Component 2 (flexible component) allows for country specificity and comprises 7 measures that will be selected for each country based on the specific PFM issues in that country that are being strengthened to strengthen the system overall and/or prevents development partners from using country systems in that country. The measures themselves are pre-set, to ensure consistency of measurement and credibility. Countries will select the measures and agree them with their development partners, and they may change over time.

There was broad consensus that a revised indicator should include a fixed global component, and a flexible country component. Opinions on the number and balance of measures across the two components however diverged, with some respondents favouring eight measures in both components, to allow for better coverage of the PFM cycle, and others favouring more measures in the flexible component.

Further discussion on the number of measures in each component, and the coverage of measures in the fixed component will be required

### Measurement methodology

It is proposed that indicators from the Public Expenditure and Financial Accountability (PEFA) framework are used. This means that for any one country that participates in the GPMF, the PFM indicator will comprise 7 fixed PEFA indicators, which are the same for all countries, and 7 flexible PEFA indicators that are country-specific.

The use of PEFA is proposed because:

- It is the instrument commonly used at country level to assess risk by almost all development partners, even if many undertake additional assessments. Harmonised risk assessment is one of the main purposes of the framework. It is also a well-established methodology, with significant credibility, and is applied in all countries participating in the GPMF;
- It is the instrument commonly used to diagnose country PFM systems and prioritise which PFM should be strengthened for improved public financial governance and service delivery;
- Also, using PEFA means that a trusted scoring methodology and existing data are used.

However, while the GPMF is applied every second year, PEFA assessments are normally 4 to 5 years apart, on the logic that PFM improvements take time. A key choice would be whether an update will be done for the 14 indicators used specifically for the GPMF, if the latest PEFA assessment was done more than a set number of years before the year assessed for the rest of the GPMF.

A second complication is the update of the PEFA framework, which is currently still under discussion. The update proposals include both new measures and adjustments to existing measures. This means that there will be a transition period, in which some countries' PFM indicator will use the previous framework and others the new framework. This can be managed in presentation, as long as the core fixed indicators are in both frameworks, even if their scope and/or measurement changed.

PEFAs are also not always public documents. However, as the GPMF is based on voluntary participation by countries, the assumption is that participating countries would not object the use of selected indicator results for GPMF purposes.

Most respondents agreed on the use of PEFA, and the need to update the PEFA measures selected only if the existing PEFA Assessment is not sufficiently recent. Particularly development partners however, raised concerns about the cost and credibility of processes for updating in between PEFA Assessments for GPMF purposes. A solution may be to set the parameters of when an update is required broadly, so that a GPMF-update is the exception rather than the rule. For example, if the most current PEFA Assessment predates the GPMF assessment before last, an update should be attempted. Such broad setting of the parameters would not only be on account of cost factors, but also because PFM improvements take time to reflect in measurement results<sup>2</sup>.

More than one respondent also raised the impact the selected use of PEFA indicators may have on the continued use of the framework overall, and country reform choices. This concern would apply particularly to the fixed component, as the country-component would relate to agreed country reform priorities, often arising out of a full PEFA assessment. An alternative to using the PEFA framework for the fixed component would be to continue to use the current methodology, and underlying CPIA measures for this component, but add a country-specific second component, comprising country-level selected PEFA indicators.

In conclusion, the proposal is to use selected indicators from the PEFA Framework, and the most recent PEFA assessment data. The GPMF process may necessitate an update of the indicators selected, when the most recent PEFA was used already for two previous GPMF processes. It is also possible to imagine a country and its development partners agreeing on an update even if the current PEFA is more recent than two GPMF processes ago, but an update is likely to present data that are significantly different in important aspects. In all cases the GPMF presentation would need to reflect the year in which the indicators were calculated.

## Measures

Significant further discussion will be required on the selection of PFM measures.

**Component 1:** Table 1 on page 8 sets out the proposed fixed, core measures. It is proposed that each phase of the PFM cycle, linked to the use of country systems in that dimension, is represented by one measure, except budget execution, which needs to include budget release and budget execution processes, as well as procurement measures and is therefore represented by 3 measures. This weights the fixed component towards budget execution. However, this is arguably appropriate, as using country budget planning and preparation systems carries less risk for donors using country systems, than using country disbursement, execution and accounting systems. Figure 1 below sets out the 4 key budget phase dimensions to be used, with the associated UCS dimensions<sup>3</sup>.

In addition one further measure is proposed on budget credibility<sup>4</sup>: this will only result if processes across the budget cycle are in place, and therefore can be described as an intermediate PFM

---

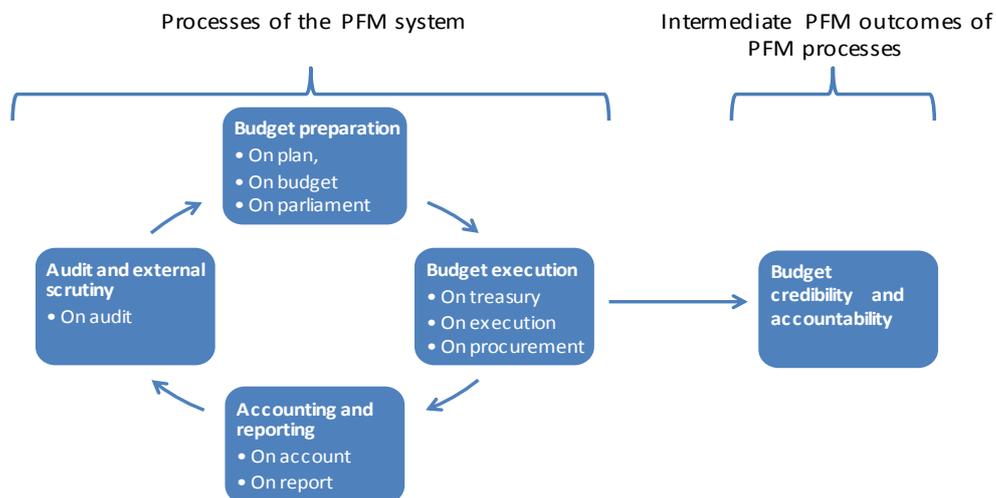
<sup>2</sup> An alternative mitigation measure would be to add the simplicity of calculating an indicator to the criteria for selecting indicators, which would affect the cost of calculation in between PEFA assessments.

<sup>3</sup> These draw on the original CABRI framework for measuring use of country systems across the budget cycle, adding on procurement and on execution, where the latter refers to the use of country internal controls.

<sup>4</sup> A second potential indicator to be used at the outcome level is indicator 14 in the 2015 Testing Draft, which measures the presence of key fiscal strategy processes as well as fiscal balance outturns compared to original forecast. This would be a key indicator for donors considering the use of country systems. The indicator is not proposed at this stage, because it would not have a data point (or the two data points required if a summary

outcome, necessary for the PFM system to result in fiscal discipline, strategic resource allocation and efficient use of resources for service delivery.

**Figure 1. Six dimensions of the PFM system for measurement**



Given the purpose of 9a and its relationship to 9b, the key criterion used to select indicators are whether performance against the PFM aspect being measured is more critical for development partners to use country systems, than alternative aspects. A second criterion is that the indicator selected must measure aspects that are likely to be critical across countries. Further aspects that are critical to any one country will be added through Component 2 of the indicator, which allows countries to select any other indicators from the PEFA framework.

This means that the framework of fixed indicators below does not measure the PFM system overall, but specific aspects of the PFM system that are more critical than others in most cases to address the risk concerns of development partners arising from a country's PFM systems. While addressing these concerns may not be sufficient for development partners in practice to use country systems, it is necessary.

**Component 2:** The proposal is that countries would select critical measures for the country aligned with targeted country PFM reforms, in agreement with development partners at country level. An option would be to require that this component mirrors the construction of Component 1. However, the proposal is to leave it open, as countries' PFM reforms may focus more in some dimensions than in others, and the requirement for an agreement between development partners and countries on the proposed measures would assist to ensure a country-appropriate spread of measures, given PFM system weaknesses and donors perceptions of risk.

indicator is presented following the methodology proposed) for the majority of countries for quite some time. The inclusion of the indicator should be considered in future though.

**Table 1. Proposed fixed (core) measures**

Dimension	Measure	Description PEFA 2011	Description Testing Draft 2015	Discussion
Budget credibility	PI 2 Composition of Expenditure outturn, compared to original approved budget	Measures extent of variance in composition of expenditure outturn over three years and expenditure charged to a contingency vote. For an A score variance in expenditure composition must exceed 5% in no more than one of the last three years, and average expenditure charged to the contingency vote is on average less than 3% of the original budget.	Proposed PI-2 in revised PEFA is refined by separating composition by economic from administrative/functional/programmatic classification, and thereby allowing deviation in interest payments to be measured, which sharpens the measure for the purposes here. A D score in any of the dimensions also now reflects any state of affairs in which the requirements for a C score or higher are not met.	Predictable budget execution is critical to give DPs confidence that their cash resources will be used as planned. This measure was chosen rather than the aggregate expenditure outturn measure, as sectors can still experience variance even if the aggregate outturns are on target.
Budget preparation (On plan, on budget and on parliament)	<p><b>2011 PEFA:</b> PI 6 Comprehensiveness of budget information included in budget documentation</p> <p><b>2015 Testing Draft:</b> PI 5 Comprehensiveness of information included in budget documentation</p>	The indicator measures how complete the executive's budget proposal is. It lists 9 information elements and an A score means that a country's budget documentation includes 7 to 9 of the listed elements.	PI 6 is PI 5 in the Testing Draft. The measure is strengthened insofar as the list of information elements has been extended to include a number of elements pertaining to the assumptions underlying the budget and fiscal risk, and separated into basic elements (deficit forecast, previous year budget outturn, current year budget, aggregate budget data in main classification heads, and a detailed breakdown of revenue and expenditure), and additional elements. For an A score all of the basic elements need to be included, and at least 6 of the additional 12 elements. A D score also now reflects any state of affairs in which the requirements for a C score or higher are not met.	A more complete budget proposal submitted to the legislature would provide assurance to DPs of budget commitments against their resources, as well as information to assess fiscal risk, particularly in the Testing Draft version of the indicator. It also provides information on the outcomes of the budget preparation process. An alternative indicator would be the overall fiscal transparency indicator (PI 10 in the 2011 Framework and PI 9 in the 2015 Testing Draft). This indicator measures public access to key budget documents across the budget cycle, but does not directly measure the completeness of these documents. The budget document completeness indicator is therefore proposed: a publicly available but significantly incomplete budget document would not provide DPs with assurance to use budget preparation systems. A second alternative would be to combine the budget completeness and availability of documentation indicators, but that would agreement on the method to calculate an average score.

Dimension	Measure	Description PEFA 2011	Description Testing Draft 2015	Discussion
Budget execution 1 (On treasury)	<p><b>2011 PEFA:</b> PI 16 Predictability in the availability of funds for commitment of expenditures</p> <p><b>2015 Testing Draft:</b> PI 21 Predictability in the availability of funds to support service delivery</p>	The indicator measures whether ministries receive reliable information on the release of funds for expenditure, specifically cash flow management practices, the reliability and horizon of in-year information to spending units and the frequency and transparency of adjustments to budget allocations. An A score means that a country prepares a cash flow forecast for the fiscal year and updates it monthly on the basis of actual cash inflows and outflows, MDAs can plan and commit expenditure for at least 6 months, and significant adjustments to the budget allocations occur only once or twice a year and are done in a transparent way.	PI 16 is comparable to PI 21 in the Testing Draft. PI 21 strengthens the 2011 PI 16 by adding a dimension on the extent and frequency of consolidation of government cash balances, and changing the scoring method so that each dimension is scored separately. An A score in the additional dimension would mean that all central government bank and cash balances or consolidated daily. A D score in any of the dimensions also now reflects any state of affairs in which the requirements for a C score or higher are not met.	The predictability of budget releases and few adjustments to budget allocations provide DPs with assurance that funds that are on treasury – i.e. disbursed through government systems – will be disbursed predictably, and that co-funding for a supported sector will also be released predictably, thereby addressing programmatic risk for donor programmes managed through government systems. The indicator also provides information on the risk that DP funds disbursed to the Treasury will be deviated given the fungibility of cash.
Budget execution 2 (On execution)	<p><b>2011 PEFA:</b> PI 20 Effectiveness of internal controls for non-salary expenditure</p> <p><b>2015 Testing Draft:</b> PI 24 Effectiveness of internal controls for non-salary expenditure</p>	The indicator measures the effectiveness of expenditure commitment controls; the comprehensiveness, relevance and understanding of other internal control rules/ procedures; and the degree of compliance with rules for processing and recording transactions. An A score would mean that commitment controls are in place and limit commitments to cash availability and approved budget allocations; other internal control rules and procedures are relevant, and incorporate a comprehensive and generally cost effective set of controls, which are widely understood; and compliance with rules is very high and any misuse of simplified and emergency procedures is insignificant.	PI 20 in the 2011 Framework is PI 24 in the Testing draft. The proposed PI 24 drops whether expenditure controls are comprehensive, relevant and understood, but adds whether expenditure management duties are segregated. For an A score in this new dimension, a country would need to have appropriate segregation of duties throughout the process, with clear responsibilities. The Testing Draft PI 24 is also calculated differently, with each dimension assessed separately. In the compliance dimension, the scoring is now on a quantitative basis. In each of the sub-dimensions a D score also now reflects any state of affairs in which the requirements for a C score or higher are not met.	Good performance against this indicator will give donors confidence that their resources will be used as intended when using country financial management systems. It measures both the quality of the controls, and adherence to them. Effective commitment controls would reduce the risk of DP resources being deviated for purposes not agreed at the level of spending agencies. The existence of other expenditure controls, e.g. checks and balances before funding is committed is critical to donor confidence, as would be a culture of adherence to rules and procedures. An alternative indicator would be PI 21 which measures the internal audit system. However, PI 20 is preferred as good performance against this indicator would imply that internal audits provide incentives for effective internal control. A second alternative is PI 18, which measures the effectiveness of payroll controls. Weak payroll management can lead to the deviation of government resources to cover salaries, leaving DP resources at risk. However, PI 20 is preferred as it measures controls that would apply to DP resources, and PI 18, if an issue in countries, can be included in Component 2 of the PFM indicator.

Dimension	Measure	Description PEFA 2011	Description Testing Draft 2015	Discussion
Budget Execution 3 (On procurement)	<p><b>2011 PEFA:</b> PI 19 Transparency, competition and complaints mechanisms in procurement</p> <p><b>2015 Testing Draft:</b> PI 23 Transparency, competition and complaints mechanisms in procurement</p>	The indicator measures transparency, comprehensiveness and competition in the legal and regulatory framework; the use of competitive procurement methods; public access to complete, reliable and timely procurement information; and the existence of an independent administrative procurement complaints system. An A score would mean that the legal framework covers all the key requirements; that contracts that are awarded by methods other than open competition are justified as required by the legal framework in all cases; that 90% of procurement by value is transparent; and that a robust complaints mechanism is in place.	PI 19 is comparable to PI 23 in the Testing Draft. PI 23 (Testing Draft) drops the legal and regulatory framework dimension, and replaces it with a dimension that measures the monitoring of the procurement system that is done. For an A score in the new dimension, a country would need to maintain databases (records) for contracts representing at least 90% of the value of procurement of goods, services and works, including data that will allow the proportion of procurement processes that were completed successfully and the difference between the original contract award and the actual cost to be computed. Analysis of this data is made available to management at least annually. A D score in all dimensions also now reflects any state of affairs in which the requirements for a C score or higher are not met.	Indicator 9b measures the degree to which donors are using country procurement systems as one of the sub-measures. It is therefore important to include PI 19, as a fixed core measure in the framework, as progress in the 9b procurement indicator would require progress against 9a.
Accounting and reporting (On account and on report)	<p><b>2011 PEFA:</b> PI 25 Quality and timeliness of financial statements</p> <p><b>2015 Testing Draft:</b> PI 28 Quality and timeliness of financial reports</p>	The indicator measures the completeness of the financial statements; their timeliness and the accounting standards used. An A score would mean that financial reports are prepared annually, and include full revenue and expenditure and financial assets and liabilities information, are submitted for external audit within 6 months of the end of the fiscal year; and accounting standards applied to all financial reports are consistent with international standards.	PI 25 is comparable to PI28 in the Testing Draft. The new indicator refers to annual financial reports. The proposed revised PI 25 is strengthened by more specific stipulation on the content of financial reports, including reconciliation with approved budget and inclusion of information on government's liabilities, guarantees and long-term obligations. It shortens the time for submission, and has more specific requirements for the quality and application of accounting standards. A D score also now reflects any state of affairs in which the requirements for a C score or higher are not met.	PI 25 is selected as good performance against this indicator – which measures a key output from the accounting and financial reporting system – would give DPs confidence to use these systems. The indicator measures both the content, underlying accounting standards and compliance with country requirements of financial reporting system. The revised indicator tests whether information is available that will help DPs assess fiscal risk. Progress against this indicator would be necessary for progress on the financial reporting sub-dimension of indicator 9b.

Dimension	Measure	Description PEFA 2011	Description Testing Draft 2015	Discussion
Audit and oversight (On audit)	<p><b>2011 PEFA:</b> PI 26 Scope, nature and follow up of external audit</p> <p><b>2015 Testing Draft:</b> PI 29 SAI Independence and external audit of the government's annual financial reports</p>	This indicator measures the scope/nature of audit performed (incl. adherence to auditing standards); the timeliness of submission of audit reports to legislature; and evidence of follow up on audit recommendations. An A score would mean a comprehensive audit mandate in line with audit standards; timely submission of audit reports and evidence of follow up.	PI 26 in the 2011 Framework is comparable to PI 29 in the Testing Draft. The proposed PI 29 reformulates the statement of dimensions and scoring statements, but is comparable, except in the following: PI 29 is strengthened through the inclusion of a dimension measuring the independence of the Supreme Audit Institution; it is more realistic in the time allowed for SAIs to submit audit reports, as well as a more specific statement on audit standards. A D score also now reflects any state of affairs in which the requirements for a C score or higher are not met.	Good performance against this indicator would give DPs confidence that vertical accountability loop is closed through PFM systems for the use of country systems in all dimensions. Furthermore, it would be key for donors to use country audit systems: progress against this indicator would be necessary for progress against the use of audit systems sub-dimension of 9b in the GPMF.

### Scoring

The PEFA framework scoring is based on a scale of A (high performance against the indicator) to D (a minimum set of criteria met (PEFA 2011) or not met (for most indicators in the 2015 Testing Draft) calibrated specifically for each question. A key choice for the revised indicator 9a would be whether to summarise the PEFA scores for the indicators used, or to reflect only the PEFA scores themselves.

#### *Reflecting the PEFA scores themselves*

Using the PEFA scores as is would avoid criticism by stakeholders of the methodology used to summarise the scores, as well as the risk of losing important detail in the reflection and interpretation of PFM systems strength. On the other hand however, reflecting and interpreting 14 measures for one sub-indicator in an overall GPMF in which most indicators are single quantitative measures, would also present difficulties. This argument can be taken one step further: the co-indicator for 9a is 9b, which is a single quantitative measure. If the GPMF is to measure progress against commitments made (development partners will use strengthened country systems), it requires a comparison of 9a and 9b. However, no interpretable comparison of a single quantitative score is possible with 14 ordinal scores, some of which will improve and some of which will deteriorate.

#### *Calculating a summary measure*

Should a summary measure be used, two further choices become apparent, namely whether the original PEFA scores should also be reflected in a GPMF results framework, and secondly, whether one summary measure (combining component 1 and 2) or two measures should be calculated. These questions are considered below, after a discussion of options for summarising the PEFA scores.

The first summarising option would be to use the methodology that converts the PEFA ordinal scale to an interval scale for the framework (as has been done in a number of studies on PFM system progress), and aggregate the numbers to a single, composite score. However, a number of methodological problems with this methodology have been noted over time, including that it incorrectly assumes equivalence in between notches between and within indicators and that an interval scale has a zero point – a D score in PEFA would in almost if not all cases not equate to a zero point (e.g. no system present or 100% budget variance), in either the 2011 Framework or the 2105 Testing Draft.

However, considering that the GPMF is interested in measuring progress, further options arise to calculate a summary score of the change in strength of country systems as measured by PEFA. In other words, a summary indicator that reflects the change in systems on average, complies with the requirement for the GPMF that 9a should measure the strengthening of the systems, rather than the systems themselves (which is a purpose of PEFA itself), or at least balance these aspects. There are two options for summarising the selected PEFA indicators' scores to provide a summary indicator of the change that has occurred, without assuming equivalency between scores, or that a D score is a zero point. While the proposals therefore avoids critical issues with summarising PEFA scores, for both proposals a first application would require either higher effort in countries where only one PEFA has been done to establish two data points, or the GPMF would not reflect a summary score for such countries.

**Proposal 1:** In this proposal the methodology comprises two parts to scoring the global fixed indicators.

- The **first component** would measure whether a country system is adequate for donors to use it. This would mean that using the PEFA criteria for each indicator, partner countries and development partners agree at the global level the score level that is adequate for partners to use country systems. Once this level is determined, for example C in an indicator, anything equal to or above this level would score a 1, while any PEFA score below the level would earn a 0 in the GPMF scoring. A simple scale is proposed where a 1 would be earned if there is improvement (from 0 to 1), a 0 if no improvement (a 0 remains a 0 and a 1 remains a 1) and -1 if regression (from 1 to 0). While this is a blunt scoring instrument (a country that improves 4 notches on the PEFA scale rather than 2 would not score better), any other option would again assume equivalency between and within scores.
- The **second component** would apply to the indicators targeted for improvement in the flexible component, and would just measure whether the country has succeeded in making progress against the weaknesses in this area. The interest is therefore in the direction of change, rather than reaching a minimum threshold. If no change occurs, it would score a 0; if improvement occurs a 1; and if regression occurs, a -1.  
An alternative for the second component would be to negotiate a minimum adequate level of performance at the country level for each of the 7 country PEFA scores selected. However, this will put an even higher burden on country processes. Also, the proposal as reflected would balance reaching a minimum threshold, with acknowledging direction of change.

Another way of achieving a better balance between adequacy of systems and direction of change, would be to assign and average two numerical scores to each indicator in both components -- one calculated using the adequacy method and the other the direction of change method. While there is merit in this proposal insofar as it balances adequacy and direction of change, it is not put forward as the preferred proposal so as to not overload country processes for component 2, and because for the global indicators, some countries may score as A already and would be prejudiced.

If one summary score overall is implemented, it is proposed to average the scores in both components and then add them together. This will represent an equal weighting of the two components, but would mean that all countries will score between 0 and 2, requiring decimals to be reflected to measure change. An alternative is to sum the sum of the scores for the first component and the sum of scores for the second component. This will however potentially favour the second component, as a PEFA notch change in one of the indicators of the first component may not be sufficient to reach the '1' (equal to or higher than the agreed adequate PEFA rating), whereas for an indicator in the second component, a one or half a notch change would be sufficient for a 1 score.

**Proposal 2:** This proposal assumes that negotiating a minimum level of achievement across countries would be considered too onerous, and therefore either the highest performance in terms of PEFA or change should be measured in a similar way across both components of the proposed 9a GPMF indicator. In this case either an A or positive change in an indicator would earn a score of 1, no change in a score below A a 0, or any regression a -1. In this case the summary score for a country would then be the sum of the scores for each indicator.

### *Other issues*

**Reflection of PEFA scores with a summary score:** The proposal is that the GPMF Report will for both proposals also reflect the PEFA ordinal scale scores against each indicator (to assess the current

level) and the change score of 0, 1 or -1. This additional information would mitigate somewhat the bluntness of the proposed scoring.

**Summarising the two components:** The proposal is to summarise the two components into one score. This will not only be more directly comparable with 9b, but will also ensure that the two components get equal attention. If the two components are reflected separately only, there is a risk that more analysis would be done of the first component, and that a summary score would not adequately balance adequacy and direction of change.

In conclusion on scoring, the proposal is to calculate a summary score for each component, following one of two methods. This summary score should be presented with the underlying PEFA ordinal scores. The two components should be summarised into one score, with the method depending on the overall method followed. Most consultation respondents agreed that there should be a summary score – with the underlying scores reflected -- and favoured method 1. Some respondents highlighted the need to reflect the discussion in the PEFA Assessment report for the indicators used, in country GPMF reports. Overall, feedback from the EIP consultation processes suggests that whether to summarise the selected PEFA indicators and how, require significant further discussion.

### Process

The use of a flexible indicator will require a country process to agree which indicators are selected. This would need to occur at least two years in advance of the first measurement of the revised 9a, or alternatively the first measurement of a revised 9a would only measure component 1, and include a process to measure component 2. The agreement should be reached between the country and all development partners either agreeing to increase their use of country systems, or supporting PFM reforms. Where possible existing PFM / use of country systems coordination structures should be used to agree the country-specific, flexible component.

### Conclusion

This Policy Brief has set out the challenges of and a set of proposals for revising indicator 9a of the GPMF. In summary the proposal is to construct an indicator with two components using PEFA data. The global component will comprise a set of about 7 indicators measuring country systems across the budget process, favouring the phases that carry high risk for donors using country systems. A draft set of indicators for the global component is proposed. The country component will comprise a second set of 7 PEFA indicators, which are determined at country level. The spread of this set across the budget cycle will be determined by the areas of focus for improvement in country systems specific to each country. Countries and their development partners will agree on the 7 indicators that will be measured. The proposal is to summarise the two components, following one of two methods, but to reflect the original PEFA ordinal scores as well.

The consultations to date have identified the following as important points for further discussion:

- The use of PEFA in this way, particularly for the global component of indicator 9a;
- Whether to require or allow updates of the data for the selected indicators for GPFM purposes in between official PEFA assessments;
- The number across and distribution of PEFA indicators between the global and country components;
- The PEFA indicators selected;
- Whether the indicators should be summarised, and if so, by which method.