



Evaluating Budget Support Methodological Approach



Acknowledgements: This document was developed by the Budget Support Evaluation Steering Group of the OECD DAC Network on Development Evaluation with the aim of encouraging more effective evaluation of budget support operations. The Steering Group was led by the Evaluation Unit, Directorate General for Development and Co-operation of the European Commission and financed with a generous grant from the Policy and Operations Evaluation Department of the Ministry of Foreign Affairs of the Netherlands. Mr. Martin van der Linde (Ecorys) and Ms. Camilla Valmarana (DRN) drafted the text.



Cover Photo: School children during a visit of the budget support evaluation team in Chadiza, Zambia. (Antonie de Kemp, IOB, 2010)

Table of contents

Table of contents	1
Abbreviations	2
Summary / preliminary remarks	3
1 Introduction and context	5
2 Methodological approach for evaluating budget support	8
3 Application of the Comprehensive Evaluation Framework and the Three Step Approach	22
4 Organisational set-up of a budget support evaluation	29
Annex A	39
Tools for Step Two of the Budget Support Evaluation	39
Annex B: Tools for the verification phase	41
Annex C: Example set of key Evaluation Questions (EQs)	42
Annex D: Summary tables of the Tunisia and Mali budget support evaluations	44

Abbreviations

BS	Budget support
CEF	Comprehensive Evaluation Framework
DAC	Development Assistance Committee
EC	European Commission
EU	European Union
GBS	General Budget Support
GNP	Gross National Product
MoU	Memorandum of Understanding
ODA	Official Development Assistance
OECD	Organisation for Economic Co-operation and Development
PAF	Performance Assessment Framework
PFM	Public Finance Management
PRSP	Poverty Reduction Strategy Paper
SBS	Sector Budget Support
SMART	Specific, measurable, achievable, relevant and time bound
SWAP	Sector Wide Approach
TA	Technical Assistance
ToR	Terms of Reference
WAEMU	West African Economic and Monetary Union

Summary / preliminary remarks

This document presents a methodological framework and approach for evaluating the contribution of budget support (BS) programmes to targeted development outcomes and impact. Such evaluations aim at *‘assessing to what extent and under which circumstances budget support has successfully enhanced the policies, strategies and spending actions of the partner government so as to achieve sustainable national and/or sector level development outcomes and a positive impact on poverty reduction and economic growth’*.

The proposed methodological framework explicitly recognises that budget support is not a development programme or policy in itself, but is rather a means to support a set of policies and public spending actions of a partner country government in order to achieve the final objectives as regards economic growth and poverty reduction. The key distinctive features of the proposed methodology are:

- the **Comprehensive Evaluation Framework (CEF)** which sets out the hypothesised sequence of effects of BUDGET SUPPORT programmes across five analytical levels (budget support inputs, direct outputs, induced outputs, outcomes and impact) included in – and interacting with – the overall national context within which budget support is provided, and;
- the **Three Step Approach**, whereby: i) Step One encompasses the assessment of the inputs, direct outputs and induced outputs of budget support (levels 1, 2 and 3 of the CEF) including the analysis of the causal relations between these three levels ; ii) Step Two encompasses the assessment of the outcomes and impact of the government's policies, strategies and spending actions, which donors supported and promoted with budget support, and identification of the main determining factors of those outcomes and impact (levels 4 and 5 of the CEF), through policy impact evaluation techniques; and iii) Step Three entails an exploration of the contribution of budget support to the government's policies, strategies and spending actions, which have produced the outcomes and impact identified in Step Two, to be carried out by combining and comparing the results of Steps One and Two.

The document provides a conceptual and overall methodological approach for structuring and carrying out the evaluation of a budget support programme at country-level. It provides information and insights to guide the evaluation teams when applying the CEF and the Three Step Approach, while at the same time pointing out that the reconstruction of the specific Intervention Logic(s) of the budget support programme(s), the formulation of evaluation questions and the selection of tools to be used especially in Step Two, will need to reflect: i) the specific country context within which the budget support programme(s) has been implemented; ii) the specific characteristics and scope of the budget support programmes being evaluated; iii) the availability of data and studies; and iv) the budget and time available to the team for analytical activities; V) the objective of and intended use of the evaluation.

At the same time, the document points out that an evaluation should not be seen as a stand-alone activity which is carried out every 5 to 7 years. The process of preparing an evaluation and a monitoring framework should start at the very beginning of a budget support programme by clearly defining the objectives, strategies and performance indicators of the programme, collecting base line data, identifying back-ground studies which need to be carried out during the implementation of the budget support programme, etc. It is also crucial that the policy dialogue and the monitoring of the Performance Assessment Framework (PAF) are well documented and archived throughout implementation.

The timing and scope of an evaluation can be defined in the Memorandum of Understanding (MoU) of a joint budget support programme /and/or in the financing agreements signed with the various donors. The timing of a budget support evaluation should be fine-tuned with the budget support-monitoring arrangements and the

implementation and finalisation of other evaluations, studies and surveys expected to provide important inputs for the budget support evaluation and for other complex evaluations. Budget support evaluations require considerable time and resources. Though specific requirements will depend on the evaluation scope, purpose and management, the pilot evaluation experiences imply that a 12-18 month timeframe will likely be required. As evaluability and data collection are improved over time, the resources required for each evaluation should decrease.

The three steps of a budget support evaluation could either be carried out in one limited (demarcated) period or separately at different times with the obvious caveat that Steps One and Two must precede Step Three and that the three steps cover the same period of evaluation. Carrying out the steps at different times does not reduce the overall evaluation effort, but has the advantage of avoiding the concentration of all efforts in a limited period of time under one contract with the ensuing complex organisation and coordination processes. On the other hand, separating the execution of the three steps has the disadvantage of not making use of a possible synergy and cross-fertilisation when Steps One and Two are carried simultaneously, while it also creates a risk of less coherence between the analyses of the three steps.

A BS evaluation covering all three steps of the CEF will include the following six phases: (i) preparation, (ii) inception, (iii) data collection and preliminary analyses, (iv) in-depth analyses and validation, (v) synthesis and (vi) dissemination of results.

Finally, it is reminded that this document presents an updated version of the approach developed in 2007/2008 for the evaluation of BS programmes under the auspices of the OECD/DAC Network on Development Evaluation. The authors, while accepting full responsibility for this document, are indebted to the authors of the original document as well as to the members of the OECD/DAC Steering Group for the evaluation of budget support for the comments and guidance provided.

+++++

1 Introduction and context

1.1. Context

Providing aid to developing countries in the form of budget support (BS) has become more and more prominent since the start of this century. Both donors and partner countries¹ felt that this aid modality would enhance the effectiveness and efficiency of development cooperation and would contribute to more sustainable development. The increased use of this (new) aid modality sparked off a debate about how to evaluate its effectiveness and efficiency. It was felt that a common understanding for how to approach budget support evaluations was needed in order:

- to upgrade and ensure the quality of BS evaluations;
- to promote comparability of results and the exchange of lessons learned between all stakeholders in different countries, and;
- to select and design appropriate tools for BS evaluations.

For that purpose, the Evaluation Unit of the Europe Aid Co-operation Office of the European Commission took the lead in 2007 - under the auspices of the OECD/DAC EVALNET – a Steering Group for the evaluation of budget support. One of its first activities was commissioning the development of a common approach for the evaluation of budget support², followed by two other methodological papers.³

So far, the new approach has been applied and tested in three evaluations of budget support, namely in:

- Tunisia, an evaluation commissioned by the EC (2011);⁴
- Mali, an evaluation commissioned by the EC, Canada and Belgium (2011)⁵, and;
- Zambia, an evaluation commissioned by Germany, the Netherlands and Sweden (2011).⁶

These evaluations have resulted in three elaborated evaluation reports, two synthesis notes and a Joint Methodological Note reflecting on the experiences with applying the new BS evaluation approach⁷.

1.2. Purpose and structure of this document

This document is meant to present in a concise form, the methodological approach for evaluating BS programmes, as outlined in the documents mentioned in section 1.1, so that there will be one single reference document as regards the methodological approach of BS evaluations. Experiences of the three pilot BS

¹ In accordance with generally accepted terminology, the terms “partner country” and “partner Government” are used in this document for countries and governments receiving (being the beneficiary of) budget support.

² Caputo, E., Lawson, A. and van der Linde, M. (2008), *Methodology for Evaluation of Budget Support Operations at Country Level*, Issue paper, DRN-ADE-EC-NCG-ECORYS.
http://ec.europa.eu/europeaid/how/evaluation/evaluation_reports/2008/budget_support_en.htm.

³ Compennolle, P. and de Kemp, A. (2009), *Tools for ‘Step 2’: The evaluation of the impact of government strategies and a Paper on ‘Methodological details’*.
http://ec.europa.eu/europeaid/how/evaluation/evaluation_reports/2008/budget_support_en.htm.

⁴ Caputo, E., Valmarana, C., Mouley, S., Corm, G. avec la collaboration du cabinet BADIS (2011), *Evaluation des opérations d'aide budgétaire de la Commission Européenne à la Tunisie entre 1996 et 2008*, Brussels, DRN.

⁵ Lawson, A., Habas, J., Keita, M., Paul, E., Versailles B. and Murray-Zmijewski A. (2011), *Evaluation conjointe des opérations d'aide budgétaire au Mali 2003-2009*, Brussels, ECO Consult.

⁶ De Kemp, A., Faust, J. and Leiderer, S. (2011), *Between high expectations and reality: an evaluation of budget support in Zambia*, Bonn/the Hague/Stockholm, BMZ / Ministry of Foreign Affairs / Sida.

⁷ These documents, except the Joint Methodological Note, are available on:
http://www.oecd.org/document/61/0,3746,en_21571361_34047972_33637693_1_1_1_1,00.html

evaluations implemented in Tunisia, Mali and Zambia have been merged with the original methodological documents published in 2008 and 2009. At the same time possible inconsistencies between the above mentioned documents have been removed, while the approach has also been streamlined in order to make it more feasible to conduct BS evaluations in a reasonable time frame.

The definition of BS and the objective of BS evaluations are presented in the two following sections of this first chapter. Chapter two is devoted to a presentation of (i) the Comprehensive Evaluation Framework (CEF), based on the intervention logic on which BS programmes are generally founded (implicitly or explicitly), and (ii) the Three Step Approach for applying the CEF. Recommendations and suggestions on how to apply the CEF and the Three Step Approach in a particular country context are presented and elaborated in chapter three. Chapter four contains descriptions, recommendations and suggestions as regards the organisation of the evaluation process. Further information on managing joint evaluation processes can be found in the OECD's *Guidance on Managing Joint evaluation* (OECD, 2007)

1.3. Definition of budget support

In 2006, the OECD/DAC has defined budget support as follows: *“budget support is a method of financing a partner country’s budget through a transfer of resources from an external financing agency to the partner government’s national treasury. The funds thus transferred are managed in accordance with the recipient’s budgetary procedures”*.⁸ Budget support is subject to specific disbursement conditions, as defined in the financing agreements. The provision of BS is accompanied by a policy dialogue focused on the main policy and reform issues of the partner government, and sometimes also by capacity development assistance. Determination and assessment of disbursement conditions are part of the policy dialogue.

BS may be provided either as General Budget Support (GBS) or as Sector Budget Support (SBS). GBS is meant to support the implementation of a national development strategy. In that case, the policy dialogue, disbursement conditions and capacity development assistance are focused on the overall policy objectives and budget priorities of the partner country. SBS is meant to support the implementation of a sector development programme and the policy dialogue, disbursement conditions and capacity development assistance are thus focused on sector specific policy issues and budget allocation priorities.

Many donors link the provision of BS also to a commitment of the partner government to adhere to good governance principles, including the respect of human rights.

1.4. Objective of a budget support evaluation

Drawing on the above definition, the objective of a budget support evaluation, reads as follows: *to assess to what extent and under which circumstances the budget support has enhanced the policies, strategies and spending actions of the partner government so as to achieve sustainable national and/or sector level development outcomes and have a positive impact on poverty reduction and sustainable and inclusive economic growth*.⁹ In this context it is not excluded that the assessment will identify unintended, possibly negative, effects of budget support operations.

Usually the stakeholders (donors and the partner government) request the evaluators to include also a forward looking component, which should generate lessons learnt and recommendations as regards:

⁸ See OECD/DAC, 2006, *“Harmonizing donor practices for effective aid delivery, volume 2*.

⁹ This objective incorporates the five standard OECD/DAC evaluation criteria: relevance, efficiency, effectiveness, impact and sustainability.

- the conditions under which budget support could have the envisaged effects, the risks of negative effects and the possible intensity and nature of such positive and negative effects;
- the design and implementation of future budget support programmes in the country concerned;
- changes in government policies, institutional structures and administrative arrangements in the partner country concerned, which might increase the overall effectiveness and impact of public policy and spending actions, and consequently also of budget support.

2 Methodological approach for evaluating budget support

2.1. Introduction

The methodological approach for evaluating BS is presented in this chapter. Section 2.2 contains a presentation of the Comprehensive Evaluation Framework (CEF) with its five levels referring to respectively: BS inputs, direct outputs, induced outputs, outcomes and impact. The Three Step Approach explaining how the CEF has to be used and analysed, is presented in section 2.3. The other three sections of this chapter are devoted to respectively: the key evaluation questions, adjusting the focus of the evaluation to the type of BS and the expected outputs of a BS evaluation.

2.2. The Comprehensive Evaluation Framework

2.2.1. Structure of the CEF

The proposed comprehensive framework for the evaluation of budget support programmes recognises the fact that budget support is a contribution to the implementation of the policy and public spending actions of a partner government. In the Comprehensive Evaluation Framework (CEF) a distinction is therefore made between (a) the direct outputs of the BS programmes, (b) their effects on government's policies and spending (induced outputs), and (c) the effects of the latter on the outcomes and impact targeted by BS.

More precisely, the CEF¹⁰ consist of five different levels (see figure 1), namely:

Level 1: Budget support inputs

Level 2: Direct outputs of budget support: improvements in the relationships between external assistance and the national budget and policy processes.

Level 3: Induced outputs: expected positive changes in the quality of public policies, the strength of public sector institutions, the quality of public spending (increased allocative and operational efficiency), and consequent improvements in public service delivery.¹¹

Level 4: Outcomes: envisaged positive effects at the level of final beneficiaries – service users and economic actors – due to improved government policy management and service delivery.

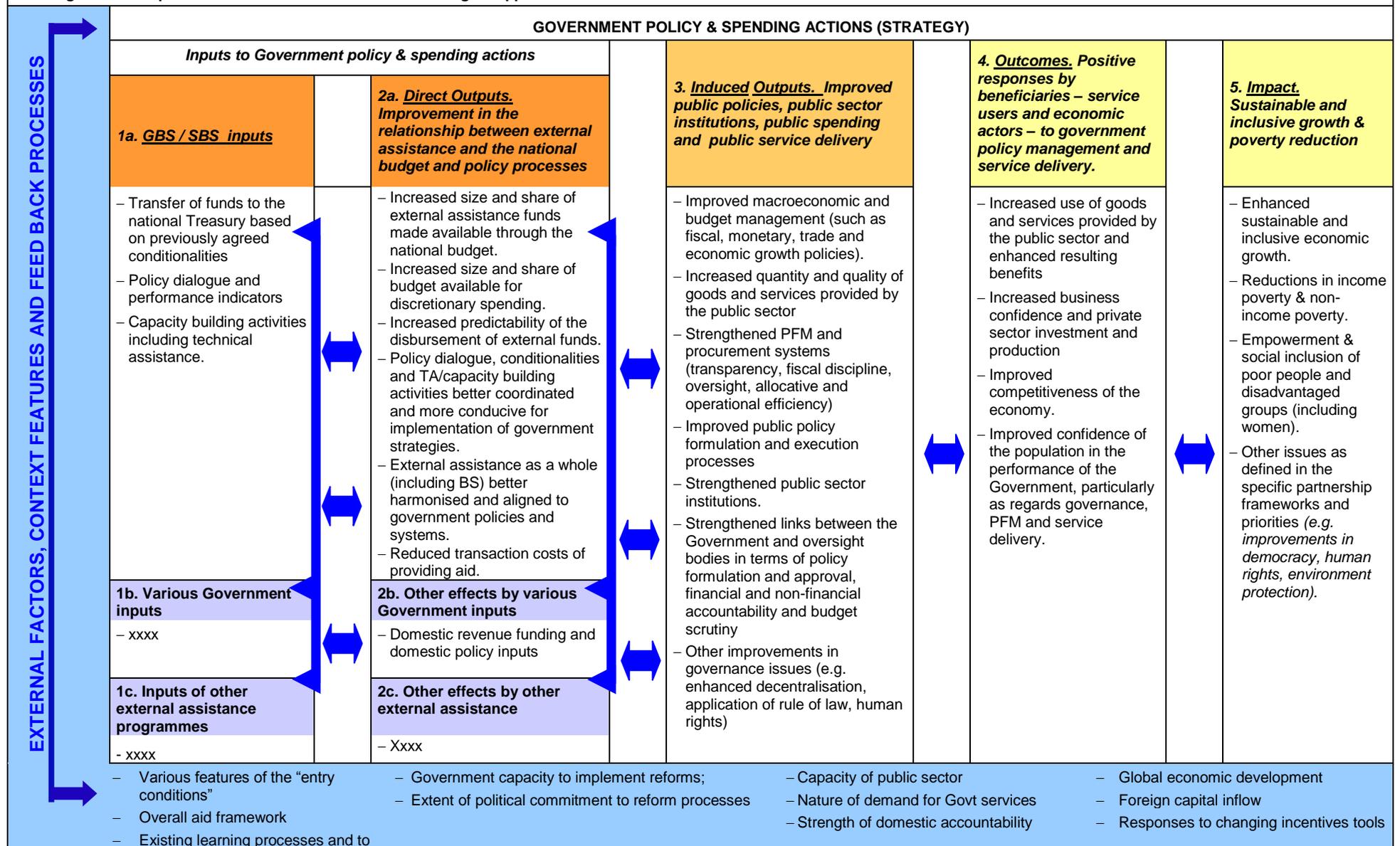
Level 5: Impact: envisaged positive effects on sustainable economic growth, poverty reduction, empowerment of the poor and improvements in their real incomes, and other issues and priorities specified in the BS programme (s) being subject of the evaluation.

The CEF includes also a final row encompassing all columns referring to the context in which BS is provided. That final row highlights that the context is not only significant for the design of the budget support programme, but also affects and is affected-by the whole sequence of effects across the evaluation framework.

¹⁰ The CEF presented in this document is the off-spring of earlier work on developing a suitable approach/methodology for BS evaluations, namely: (i) Booth, D and A. Lawson (2004), *Evaluation framework for General Budget Support*, (ii) IDD and associates (2006) *Joint evaluation of General Budget Support 1994-2004*, and (iii) Caputo, E, A. Lawson and M. van der Linde (2008), *Methodology for evaluations of Budget Support Operations at country level, issue paper*.

¹¹ According to the current DAC definitions, the accomplishment of a policy reform has to be considered as an "output" in the intervention logic. It cannot be considered as an "outcome" because it does not represent per se a benefit to the people targeted by the BS. On the other hand, the accomplishment of a policy reform is not a direct output of the BS programme although the programme may have been designed to promote it, but rather an accomplishment of national stakeholders influenced by a number of other factors including BS. That is why this crucial level of the CEF is called "induced outputs".

Figure 1 - Comprehensive Evaluation Framework for budget support evaluations



EXTERNAL FACTORS, CONTEXT FEATURES AND FEED BACK PROCESSES

This CEF is based on the intervention logic – or programme theory – on which a budget support programme itself is usually founded and sets out the hypothesised sequence of effects of BS across the five levels¹². Analysing and evaluating the logic and the underlying assumptions, budget support makes it possible to unpack the BS processes and to identify the causal relationships between inputs, direct and induced outputs, outcomes and impact, while taking into account the influence of exogenous factors. However, these relationships are not always traceable and/or unequivocal, because BS is not a development programme in itself, but an aid modality that supports the development strategy of a partner country. This has two important consequences for the evaluation of BS and for the CEF, namely:

- some of the outputs of BS are directly and to a large extent solely related with the provision of BS (level 2, direct outputs) while other outputs of BS are not directly and/or not solely related with BS (level 3, induced outputs). The induced outputs refer particularly to changes (improvements) of government policies and services, which are not the result of the BS programme alone but also of various other government actions and decisions, other external assistance programmes (see the cells 1b, 1c, 2b and 2c of the CEF); and external non-government related factors.
- the evaluation will have to encompass a broad assessment of (i) the outcomes and impact as regards social and economic development which were targeted by the government policies supported by the BS programmes, and (ii) the factors which have contributed to achieving those outcomes and impact (government policies, aid programmes, domestic and international factors outside the control of the government, context features mentioned in the bottom line of the CEF, etc.). This assessment will make it possible to identify the contribution of BS to the achievement of development results at the outcome and impact levels. The Three Step Approach, proposed in section 2.3, will allow this type of analysis.

Details on the content of each of the 5 levels of the CEF and issues to be analysed are summarised in the following sub sections.

2.2.2. Level one: the budget support inputs

Budget support consists of a package of inputs usually provided by a number of jointly operating donors.

Main inputs would typically include:

- a **transfer of funds** to the Consolidated Fund of the National Treasury of the partner country, disbursed on the basis of previously agreed disbursement conditions;
- a **policy dialogue** as regards major policy and reform issues and performance indicators, focused on either the formulation and implementation of a national development strategy (in case of GBS) or on a sector development programme (in case of SBS);
- provision of **capacity building support**, including technical assistance (TA) aimed at strengthening the capacity of the partner government in areas related to the objectives of the BS programme. Support for the implementation of a Public Finance Management Reform Programme is often an important component of these capacity building activities.¹³

¹² In case there is more than one BS programme in a country, it is recommended to make a CEF for each BS programme: one for the joint GBS programme and one for each joint SBS programme. In such a case it might be useful to make also one synthetic CEF with a view to set out the overall sequence of expected effects. It is not recommended to make a CEF for each BS financing agreement (signed by one single BS donor and the partner government) .

¹³ In some cases, capacity building programmes are part of the BS programme, in other cases they are provided under separate financing agreements. There are also cases of PFM reform programmes which are funded by the national budget of the partner country. Irrespective of the source and modalities of financing, capacity development programmes should be part of the BS evaluation as long as they are clearly linked to the BS programme.

At the first level of the CEF, the BS evaluation should include an inventory and analysis of (i) all the inputs planned and actually provided (in this context the scope of inclusion of capacity building projects that are of relevance but not part of BS programmes need to be defined), and (ii) the disbursement conditions and other specific BS modalities which may differ from one donor to another. Furthermore, an analysis should be made whether the inputs constituted a relevant and appropriate package of support for the country concerned in view of the BS objectives, the initial socio-economic conditions in the country concerned and the policies of the partner government.

The column of level one in figure 1 refers also to various government inputs (1b) and inputs of other external assistance programmes (1c). Although these inputs are not part of the prime subject of evaluation, some of these inputs may play a role in the chain of effects considered at levels 2 and 3 and will have to be taken into account in the evaluation as well.

2.2.3. Level two: the direct outputs of budget support

Level 2 of the CEF deals with the direct outputs of the budget support programme(s), in particular as regards improvements in the relationship between external assistance and the government budget and national policy processes. It should be noted that the CEF distinguishes direct outputs and induced outputs (indirect effects of BS which are also influenced by many other factors; see level 3 of the CEF). The CEF postulates that budget support will produce the following direct outputs:

- increased size and share of external funding made available through the government budget;
- increased size and share of the government budget available for discretionary spending;
- increased predictability of the disbursements of external funds;
- improved policy dialogues and more effective (disbursement) conditions through better coordination, more consistency with the government priorities and stronger incentives for effective implementation of government strategies;
- better coordinated technical assistance and capacity building support provided in the context of the budget support programme(s), which are more consistent with government priorities and more conducive to the effective implementation of government strategies;
- a better coordinated and harmonised total package of external aid provided to the partner country, which is also more aligned with the government's policies and implementation systems;
- reduction of transaction cost of providing and receiving aid at the level of both the partner government and the donors.

The column of level two in figure 1 refers also to various effects of government outputs (2b) and effects of other external assistance programmes (2c). These effects are not part of the prime subject of evaluation, but they have to be taken into account because (some of) these effects may influence the induced outputs (level 3).

2.2.4. Level three: the induced outputs

Level 3 of the CEF refers to the induced outputs of the budget support programme(s). Induced outputs are outputs which are not directly produced by the BS inputs and the direct outputs, but require another actor (in this case the government) to produce them. The induced BS outputs are therefore not the result of budget support alone, but rather the result of a variety of government actions which may be influenced by budget support but also by other factors, including the outputs of other external assistance programmes and/or other external factors.

Expected key induced BS outputs are:

- improved macroeconomic and budget management (including revenue and expenditure policies, inflation and debt management, monetary and foreign exchange policies, trade policies, etc.);
- increased quantity and quality of goods and services provided by the public sector;
- strengthened PFM and procurement systems (including improved fiscal discipline, transparency and oversight, and enhanced allocative and operational efficiency including closer links between policies and budget allocations);
- improved public policy formulation and execution processes;
- strengthened public sector institutions;
- enhanced allocative efficiency of public expenditure, including closer links between policies and budget allocations;
- strengthened links between the government and oversight bodies in terms of policy formulation and approval, financial and non-financial accountability and budget scrutiny;
- other improvements in governance issues (e.g. improved relations between central and local governments)

2.2.5. Level four: the outcomes

Level 4 of the CEF deals with the expected outcomes at the level of the beneficiaries (the population and economic actors) as targeted by the government and supported by BS programmes. These outcomes are the result of the entire set of policies, strategies and spending actions of the government (in most cases supported by BS programmes as shown in Level 3), of the reactions/responses of the relevant stakeholders to changes in public policy making and resource allocation decisions, and of the influence of other external factors. These outcomes are thus only partly influenced by the BS provided. The causal relationship between the provided budget support and the outcomes will therefore be (strongly) diluted with other influencing factors.

In the medium term, the following outcomes are expected:

- increased use of goods and services provided by the public sector in the areas targeted by the government policies and activities supported by BS programmes, and enhanced positive outcomes thanks to increased quality and quantity of public goods
- a positive response from economic actors in terms of increased business confidence and growth of private sector investment and production ;
- improved competitiveness of the economy;
- improved confidence of the population in the performance of the government, particularly as regards governance, PFM and service delivery.

2.2.6. Level five: the impact

Impact may be expected to materialise in the long term, provided the expected outcomes are produced and key assumptions as regards external factors and growth and development processes materialise (see last row of the CEF). The changes taking place at this level are the combined effects of processes of economic growth, policies and actions of the government influenced by the budget support programme(s), underlying processes of change and trends in the domestic economy and society at large, external factors, etc.

The desired/envisaged final impact – to which BS is supposed to contribute – encompasses:

- enhanced sustainable and inclusive economic growth;
- reduced income poverty and non-income poverty;
- empowerment and social inclusion of poor people and disadvantaged groups (including women).

There may well be other impact areas to consider, depending on the specific partnership framework and the related priorities established by the partner government and the main development partners, such as improvements in democracy, human rights, and environmental protection.

2.2.7. Main driving forces within the Intervention Logic of budget support

The evaluation framework assumes that there are two main driving forces which generate (most of) the effects at the levels 3, 4 and 5 of the CEF. These two driving forces are:

- the flow-of-funds effects resulting from the provision of the (additional) BS funds, and;
- the policy and institutional effects resulting from the interplay of BS funds and the non-financial BS inputs with domestic processes of policy-making, budget formulation and budget execution.

The flow-of-funds effects stem from the BS funds which are additional to domestic funds for financing the national budget and which are managed by the public finance management (PFM) system. The BS funds allow the Government either to increase the level of public spending or to reduce the level of borrowing or to increase government's savings¹⁴. When BS funds are disbursed early in the fiscal year, they may also reduce the costs of financing the annual budget by reducing the need for within-year borrowing to meet cash flow requirements. The additional funds may also have macroeconomic effects, in particular on economic growth (additional demand for domestically produced goods and services) and on interest and exchange rates (a reduction of government borrowing may help to control the interest and inflation rates). BS increases also the availability of foreign exchange and the demand for local currency, which may contribute to an appreciation of the exchange rate. This may have negative effects on the export and import substitution sectors, while import dependent sectors and the consumers will benefit.

Another (hypothesised) flow of funds effect concerns the increase of the volume of aid funds managed by the PFM system of the country. It is often argued that by channelling more aid funds through the PFM system, the system becomes stronger because of (i) the larger amounts managed and (ii) the external monitoring of the quality of the PFM (often in the context of a PFM strengthening programme).

Policy and institutional effects on the other hand, are the effects on policies and institutional processes which may result directly or indirectly from one or more BS inputs, including (i) the policy dialogue between the partner country government and the external development partners, (ii) the disbursement conditions and (iii) the capacity-building activities which may have been provided. Examples of policy and institutional effects include:

- improved macroeconomic and sector policy management;
- stronger PFM systems, including improved procurement processes and increased accountability;
- improved policies and budget formulation and implementation at the level of the line ministries;
- stronger links between policy priorities and budget allocation;
- improved management and supervisory systems and capacity at ministerial levels;
- strengthened monitoring and evaluation systems (including statistical systems);
- stronger role of Parliament and civil society organisations in monitoring budget formulation and implementation, and;
- strengthened supervisory role of the Ministry of Finance and the National Audit Office.

¹⁴ When BS funds are not immediately used for increased spending in the year of disbursement but for increasing savings or less borrowing, the flow of funds effects in that particular year will be limited to less debt and interest charges and possibly less inflation.

2.3. The Three Step Approach

2.3.1. The insufficiency of a direct causality approach

The CEF for BS evaluations as presented in section 2.2 is complex. It explicitly recognises that the induced outputs (level 3), the outcomes (level 4) and the impact (level 5) are the results of a variety of actors and factors, most notably government policy measures and spending actions which may or may not have been supported by BS or by other aid programmes, civil society and private sector initiatives, as well as other exogenous factors (e.g. commodity prices on the world market, external capital inflow, political (in)stability, etc.). In most cases it will be possible to identify the specific contribution of BS up to the induced outputs at level three of the CEF, but such direct tracing is mostly not possible at level four (outcomes) and level five (impact) of the CEF. Thus, carrying out an analysis assuming a direct causality between BS inputs and (induced) outputs on the one hand and outcomes and impact on the other hand will not be fruitful and in many cases even impossible. Furthermore, such an analysis could also lead to underestimating or overlooking external factors which might have played a (major) role in the achievement of outcomes and impact. The understanding of the limits of the direct causality analysis gave rise to the development of the Three Step Approach for the evaluation of BS, which is presented in the following sub-sections.

2.3.2. The Three Step Approach

The three step approach acknowledges that:

- effects of budget support can be traced up to level 3 of the CEF, while recognising that the induced BS outputs are also influenced by various other actors and factors not being part of the BS arrangement and package;
- the results of BS at the outcomes and impact levels are very difficult to trace on the basis of a direct causality analysis;
- BS aims to contribute to strengthening and expanding public sector services and investments, and its outcomes and impact are part of the outcomes and impact of the Government's policy implementation and spending;
- the outcomes and impact of Government spending on reducing poverty and stimulating economic development are determined by the quality of national strategies and policies, the government's commitment to actually pursue these policies, and the government's political and administrative capacities to use its resources efficiently and effectively to implement these policies.

Because of these facts, the application of the CEF needs a three step approach. These three steps are:

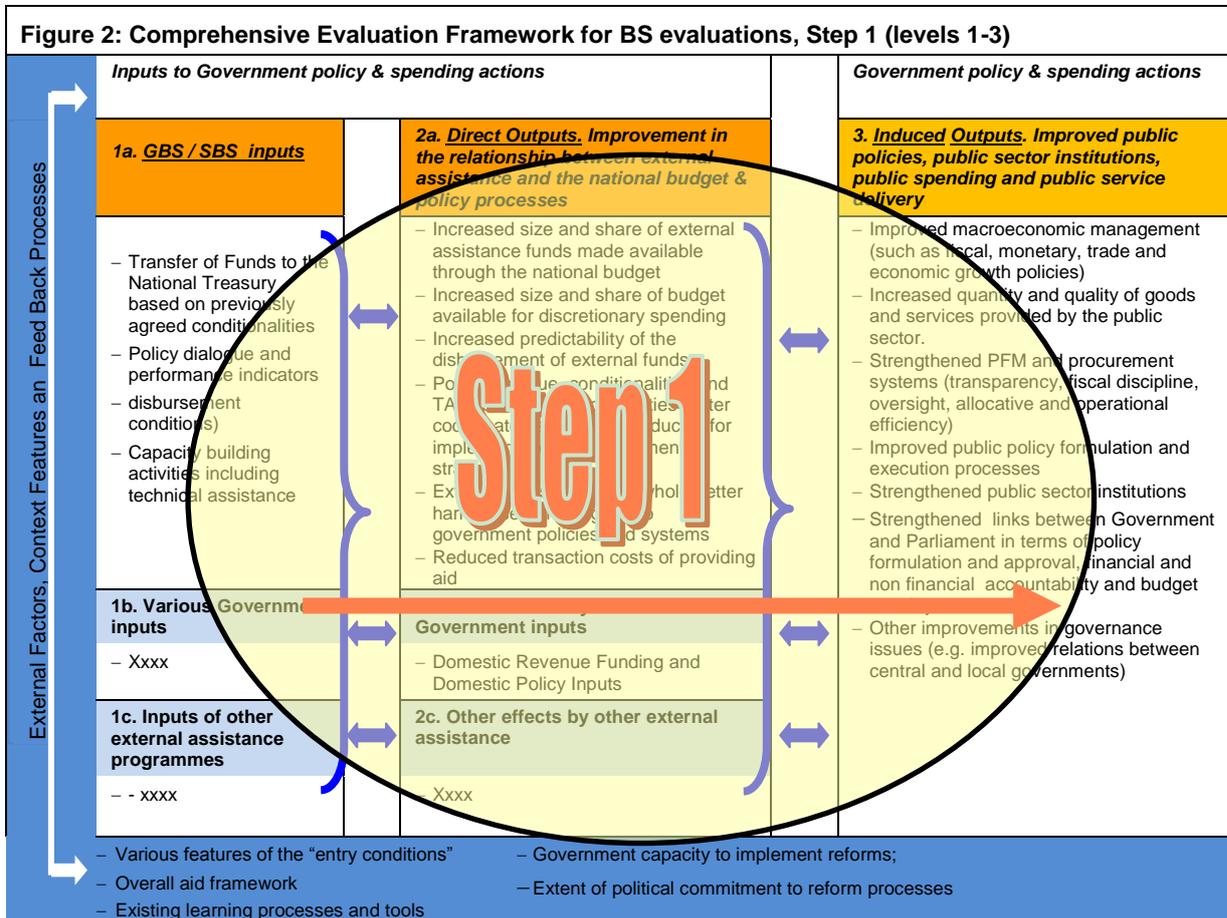
- Step 1:** Assessment of the inputs, direct outputs and induced outputs of BS (levels 1, 2 and 3 of the CEF) and analysis of the causal links between these three levels;
- Step 2:** Assessment of the expected and actual outcomes and impact as targeted by the government, which donors supported and promoted with budget support, and identification of the main determining factors of those outcomes and impact (levels 4 and 5 of the CEF);
- Step 3:** Exploration of the contribution of budget support to the government's policies, strategies and spending actions, which have produced and/or contributed to the outcomes and impact identified in step 2. This is carried out by combining and comparing the results of Steps One and Two.

It is important to note that the three steps are in principle meant as analytical steps rather than chronological steps, with the obvious caveat that Steps 1 and 2 must cover comparable periods and must

precede Step 3. Whether or not Steps 1 and 2 will be carried out in parallel or in chronological order, will depend on the organisation, data availability and requirements of a specific evaluation. Implementation in parallel has the advantage of cross-fertilisation of findings and synergy, while the chronological approach has the advantage of getting first a clear picture of the BS programmes, before the focus of step 2 is defined.

Step One

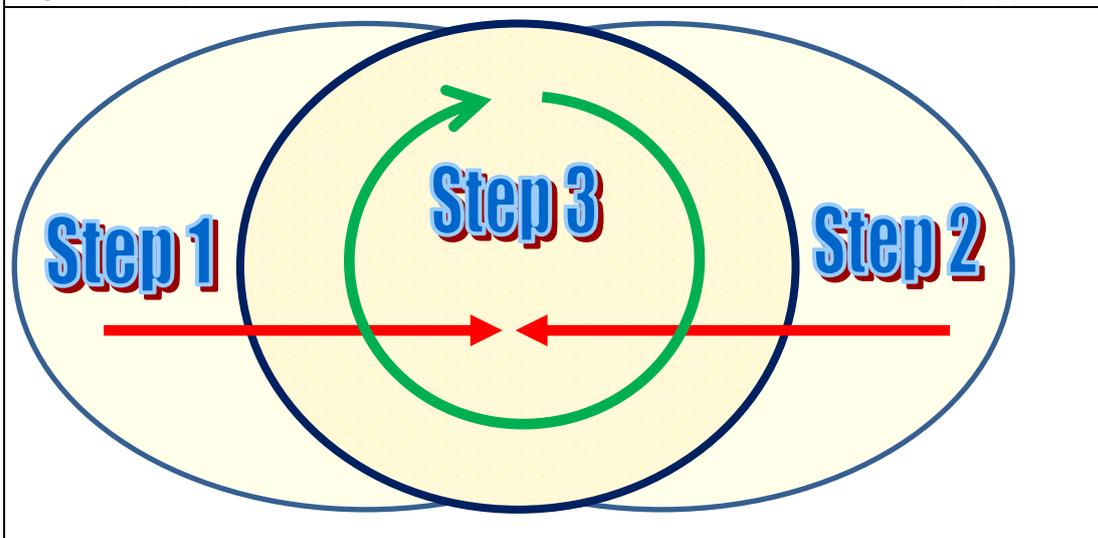
The first step of the approach consists of making a thorough inventory and analysis of the inputs, direct outputs and induced outputs of the BS (levels 1, 2 and 3 of the CEF), an evaluation of the causal relationships between the inputs and the direct and induced outputs, and an assessment of how and to what extent BS has contributed to improving the quality and adequacy of government policies and service delivery systems (induced outputs). The first step entails thus an analysis of the causal links between BS inputs at level one via direct outputs at level two to induced outputs at level 3 as shown in figure 2 (which is in fact a contribution analysis). The assessment at induced output level should be extended to include an analysis of the external factors having influenced government policies in conjunction with BS programmes.



The third step is in fact a contribution analysis. In most cases it will be difficult to quantify the BS contribution, which means that the contribution has to be assessed in a qualitative manner through logical reasoning and identifying logical linkages and contributions.

Step Three includes also (i) a comparison of the preliminary conclusions of steps one and two and (ii) verification of the coherence and consistency between the conclusions of Steps One and Two. Insights obtained from the Step Two analyses might shed new light on the preliminary findings of Step One, while drawing conclusions as regards BS contributions (Step Three) might help to specify further the conclusions of Step Two. Finally, Step Three might also allow for gaining new and/or additional insights in the role of external factors (economic, political and social developments).

Figure 4: Comprehensive Evaluation Framework for BS evaluations, the Three Step Approach



Methodological differences between the three steps

There is a significant methodological and practical difference between Step One and Step Two. The first step is basically an aid effects evaluation, while Step Two is a policy impact evaluation. This distinction calls for the use of specific and different methods for the two steps. The aid effects evaluation (Step 1) is basically an analysis of causal relations between inputs and direct and induced outputs. It assesses how BS, through its different inputs and mechanisms, has contributed to strengthening government policies, institutions, budget allocation processes, PFM and service delivery. It analyses the effects of the BS aid modality on the policy actions and institutions of the partner country and their interaction with the context, including other aid programmes and policy inputs. The policy impact assessment (Step 2) studies the outcomes and impacts targeted by the government (and the BS programmes) and the role the government policies have played in realising the actual achievements. Various methods will have to be used such as Drivers of Change Analysis, Regression Analysis and other statistical tools, (see chapter 3 and annexes A and B). Finally Step Three is again different in nature. As said before it is a qualitative contribution analysis.

Considering the importance of the external factors in both Step One and Step Two, especially as regards their contribution to the induced outputs (Step One) and their role in the determination of the outcomes and impact (Step Two), the evaluators should be aware of the fact that the Three Step approach requires a broad understanding of the social and economic development processes at national and/or sectoral level and of the political economy framework.

2.4. Evaluation questions

Because of the broad scope and complexity of budget support operations, these evaluations tend to cover a lot of material that leads to the temptation to formulate a large number of evaluation questions. Based on the pilot evaluation experience, it is advised, however, to focus the BS evaluation on a limited number (10 to 12) key evaluation questions. The questions for a specific BS evaluation have to be tailored to the specific context of the country concerned in terms of economic, social and political circumstances and to the specific focus and priorities of the BS programmes.¹⁵ It is important that there is agreement among stakeholders on the meaning and focus of key questions.

Table 1 below provides a list of key issues to be addressed by the evaluation questions, grouped per Steps and Level of the Three Step Approach and the CEF. The list is to be referred to when formulating the evaluation questions and to be used as a checklist to ensure that the selected questions cover the various aspects to be analysed. A sample set of twelve evaluation questions is provided in Annex C.

Table 1. Key issues to be addressed by the evaluation questions of a budget support evaluation

Step 1, Level 1	<p>Comparison between planned budget support inputs and those actually provided.</p> <p>Relevance and appropriateness of the design of the BS programme(s) and the mix of BS inputs in relation to:</p> <ul style="list-style-type: none"> ▪ the political, economic and social context of the partner country; ▪ the government's policy framework, and: ▪ the development partners' development assistance strategies.
Step 1, Level 2	<p>Contribution of budget support to:</p> <ul style="list-style-type: none"> ▪ increased size and share of external funding subject to the government's budgetary process; ▪ increased size and share of the government budget available for discretionary spending; ▪ improved predictability of aid flows; ▪ the establishment of an efficient and effective policy dialogue framework focused on strategic government priorities; ▪ the provision of well-coordinated technical assistance and capacity building activities focused on strategic government priorities; ▪ greater harmonisation and alignment of external assistance as a whole; ▪ reduced transaction costs of external assistance as a whole. <p>Identification of possible unintended effects.</p>
Step 1, Level 3	<p>Improvements in the areas supported through BS programmes and identification of the role played by BS in determining these changes, e.g. :</p> <ul style="list-style-type: none"> ▪ macroeconomic and budget management (revenue and expenditure policies, inflation and debt management, monetary and foreign exchange policies); ▪ quantity and quality of goods and services provided by the public sector ▪ PFM and procurement systems (fiscal discipline, enhanced allocative and operational efficiency, transparency, etc.); ▪ public policy formulation and execution processes, including strengthened public sector institutions; ▪ other governance related policies and their execution ▪ links between the government and oversight bodies in terms of policy formulation and approval, financial and non-financial accountability and budget scrutiny.

¹⁵ The collection, verification and analysis of data and information are important tasks to be carried out during the BS evaluation, but those tasks are in essence not EQs. .

	Identification of possible unintended effects.
--	--

<p>Step 2, Levels 4 & 5</p>	<p>Assessment of expected achievements in terms of development results at outcome and impact level as defined in the BS agreements, e.g.:</p> <ul style="list-style-type: none"> ▪ changes in the competitive structure of the economy (enhanced competition on the domestic market; increased competitiveness of telecommunication services, increased capacity and openness of financial services) and impact in terms of sustainable economic growth (growth of private sector investment and production,.....); ▪ changes in income and non-income poverty; ▪ changes in the use and resulting quality of public services and their impact on the livelihoods of the population: <ul style="list-style-type: none"> - for example in case of SBS for the education sector: enrolment, dropout, repetition and completion rates, gender parity, learning achievements, availability of a qualified labour force responding to market demand, employment rates, etc. - for example in case of SBS for the health sector: health centres utilisation, supervised deliveries, immunised children, ..., infant / under five / maternal mortality rates, incidence of malaria / tuberculosis / respiratory infections, etc. ▪ Changes in other key issues defined in the BS agreement, e.g. enhanced democratic governance, human rights, environmental protection, ... <p>Assessment of the extent to which the above-mentioned changes can be related to changes in macro-economic management, to PFM systems, to changes in other government policies or policy processes and / or to other external or internal factors</p>
---	--

For each evaluation question a few clear and “SMART” indicators have to be formulated, measured and assessed¹⁶. An indicator is a quantitative or qualitative factor or variable which has to be measured in order to answer the evaluation question. Standard indicators are not provided in order to allow for flexibility, which means: leaving room for adapting the BS evaluation to a specific country context. In some cases a question might be broken down into sub-questions, which implies that indicators for each sub-question will have to be defined.

2.5. Adjusting the focus of the evaluation to the type of budget support

The key issues listed in table 1 and the Evaluation Questions recommended in Annex C are likely to be relevant for both GBS and SBS. In a country where only SBS is provided, the evaluation questions will need to be applied to both the general level and to the sector or sectors for which SBS is being provided. There are two reasons for this. In the first place there are many SBS arrangements which are of a sufficient scale to generate a noticeable impact at national level, for example on the overall availability of discretionary resources within the national budget. Secondly, it will be important to identify the most significant changes taking place at national level in order to understand what is driving the changes at sector level. For example, budget allocation constraints and decisions taken at national level may have major implications for the sector being supported with BS. The same applies for a national reform programme affecting all sectors.

Furthermore, comparing national and sector level figures will put sector developments in the right perspective. For instance a modest increase in sector funding may be a significant achievement in the context of a constrained macro-economic situation and might be considered as an important contribution from the SBS arrangement. However, the assessment would be quite different in a context of fast growth in domestic revenues. Similarly, in case the quality of parliamentary scrutiny of public spending at sector level appears to have improved, that improvement cannot be attributed automatically to the SBS programme.

¹⁶ SMART = specific, measurable, achievable, relevant and time bound.

Other factors at national level may have caused that improvement or it has been the result of both national factors and the SBS programme.

Besides the need to apply most key Evaluation Questions at both aggregate and sector level, some aspects of SBS require specific consideration and emphasis. Key issues specifically related to SBS include for example the sector policy focus, sector level outputs targeted in the SBS arrangement, performance measurement at sector level, adequate funding of the sector budgets, government-donor relations at sector level, sector level alignment, and so on.

2.6. Expected outputs of a budget support evaluation

The BS evaluation should produce the following specific outputs:

- a) An inventory of the inputs provided through the BS arrangements over the period under evaluation, including the identification of possible particularities, complementarities, specific added values or trade-offs in the design and/or implementation mechanisms of the different BS programmes and BS agreements.
- b) An assessment of the performance of these BS inputs in terms of direct and induced outputs as defined in the levels two and three of the CEF (see figure 1).
- c) An identification of the changes which have occurred during the period under evaluation in relation to the outcome and impact of government policies, strategies and spending actions, and the key causal factors driving those changes (see levels 3, 4 and 5 of the CEF), including the induced outputs (level 3) related to the budget support provided.
- d) An assessment of the extent to which budget support may have contributed to the results identified at outcome and impact levels, considering both positive contributions to public policy-making and implementation processes and any (unwanted) negative side-effects which may have arisen.
- e) An assessment of the overall relevance of the budget support arrangements in the light of the evolving national and sector-specific contexts, the aid policies and the related goals.¹⁷
- f) Two sets of recommendations as regards:
 - future budget support programmes, in particular concerning the form and design of the budget support and the complementary technical assistance to be provided, and;
 - key policy and institutional issues which need to be addressed in order to increase the impact of public policy and spending actions and by implication, the effectiveness of budget support.

¹⁷ The relevance assessment includes the relevance of the instrument, the relevance of its specific design and the relevance of its actual use.

3 Application of the Comprehensive Evaluation Framework and the Three Step Approach

The Comprehensive Evaluation Framework (CEF) and the Three Step Approach should be seen as useful conceptual approaches but not as a detailed and prescriptive new methodology. The evaluation teams are advised to use: i) the CEF model when reconstructing the Intervention Logic(s) underlying the BS programme(s) to be evaluated, ii) the Three Step Approach, iii) a set of prioritised evaluation questions, and iv) a mix of methods and tools for each step, taking into account the country context, the availability of data and the amount of the evaluation budget. The following paragraphs aim at providing information and insights that could guide the teams when applying the CEF and the Three Step Approach.

3.1 Structuring the evaluation

3.1.1 Reconstructing the Intervention Logic

When applying the methodological approach described in the preceding chapter and while staying in line with the requirements set out in the ToR of the BS evaluation, the team should reconstruct the specific Intervention Logic and should propose a tailored set of evaluation questions which reflect: i) the specific context within which the BS programmes took place; and ii) the specific characteristics and scope of the budget support programmes being evaluated. The proposed five-level analytical framework, i.e. the comprehensive evaluation framework (see figure 1), will therefore need to be adjusted on the basis of a careful analysis of:

- the different budget support programmes and agreement(s), the related Performance Assessment Frameworks (PAF) and the supporting documents to which the programmes, agreements and PAFs refer;
- the various policy and strategy documents of the partner government, and;
- other documents and studies which describe and analyse the social and economic situation in the partner country.

The three tests of this approach undertaken in the period 2009-2011¹⁸, have shown that the first two levels of the CEF (i.e. the inputs and direct outputs levels) are likely to require only minor modifications to ensure that all BS components/inputs and direct outputs are adequately included, while important adjustments are likely to be required at the levels 3 to 5 of the CEF when applying it to a specific case to be evaluated. From level 3 of the CEF onwards (induced outputs), the specific objectives of the BS programme(s) and of the Government strategies will require the introduction of new or modified chains of expected results with a view to reflect the specific policy areas supported through the BS programmes as well as their intended outcomes and impact. **It is recalled that it will be useful to make a specific CEF for a joint GBS programme and one for each specific joint SBS programme, thereby facilitating the identification of the key issues of each joint BS programme to be investigated through the evaluation questions. It might also be useful to make a synthetic CEF summarising the key issues of the specific CEFs.**

¹⁸ In Tunisia, Mali and Zambia.

3.1.2 Formulating / proposing the evaluation questions

Table 1 in section 2.4 provides an overview of the key issues to be addressed by the evaluation questions. There may be a temptation of wanting to cover everything in a BS evaluation but care should be taken to keep the number of evaluation questions manageable, leading to feasible and credible answers, and also in view of the budget available for the evaluation (usually about 10 to 12 questions would be appropriate, see also the example set provided in Annex C). The final set of questions will have to reflect the key themes and the priorities shown by the specific intervention logic(s) and the key issues to be investigated as highlighted in the ToR. Thus, the evaluation questions have to be tailored to: (i) the policy content and Intervention Logic(s) of the budget support programmes to be evaluated and (ii) the political, economic and social context of the country concerned and its evolution.

In countries with more than one joint budget support programme – one joint GBS programme and one or more joint SBS programmes - the questions will need to be formulated in such a way that similarities and differences among the joint budget support programmes will be thoroughly investigated (ranging from issues related to the design of the programmes to issues related to disbursement conditions, donors' agendas and practices, policy dialogue, harmonisation among donors, etc.)¹⁹. It should be analysed whether results at sector level can be linked to the different forms of BS (GBS or SBS) and whether improvements of the quality of policy processes and policy implementation can be linked to the different foci of joint GBS programmes and joint SBS programmes.

3.2 Step One: analysis of inputs and direct and induced outputs

In Step One, the evaluators should examine the validity of the programme theory of budget support, i.e. the hypothesised cause-effect relationships as shown by the Intervention Logic of BS across the first three levels of the framework - inputs, direct outputs and induced outputs - including their relationships with the context and other possible internal or external support programmes as well as possible unintended effects. An attempt should be made to distinguish the flow of funds effects and the policy and institutional effects (see also section 2.2.7). It is acknowledged that this will not always be easy, because these effects are often intertwined and there is often interplay between them, but distinguishing – as far as possible - the effects and drivers of change and how they operate within the Intervention Logic of budget support, will help deepening the analysis of the evaluation.

When carrying out the analysis of Step One, counterfactual scenarios should be developed particularly at the level of the induced outputs, in order to gain a basic understanding of the contribution of BS inputs to the government strategy outputs that BS intended to support (*induced outputs*)²⁰. In many cases, the evaluation teams will have to rely on a hypothesised counterfactual, which may not always be easy to construct.

More precisely, Step One will encompass:

- an overview of the evolution of the context within which the BS programme(s) has (have) to be evaluated, including: (i) the overall country context within which the government strategy was conceived and implemented, (ii) the entry conditions under which provision of budget support has started, and (iii)

¹⁹ Each donor contributing to a joint BS programme has its own BS financing agreement with the partner country government. In some cases it might be worthwhile/necessary to analyse to what extent differences between these BS financing agreements have had consequences for the implementation of the joint BS programme.

²⁰ A case by case approach should be adopted, identifying and discussing possible realistic alternatives to Budget Support (projects, nothing, intervention of other international partners replacing the BS donors). Furthermore, because of the importance of the interaction of BS with other development programmes, specific care is needed to disentangle BS effects from the effects of such programmes. When adequate information is available, complementarities, synergies and other interactions on the positive and negative side should be documented.

the pre-existing framework of external aid in the country, which in most cases has been affected by the provision of BS;

- an assessment of the planned BS inputs and those actually provided, vis-à-vis the context and the scope of the BS programmes, taking into account the actual volume of funds transferred, the content and process of the policy dialogue and the disbursement conditions, technical assistance linked to the joint BS programmes, and the level of harmonisation and alignment of the BS agreements of the individual donors (see Evaluation Question (EQ) 1.1);
- an assessment of the direct outputs of the BS in terms of modified relationships between external assistance and the national budget and policy processes. Emphasis should be placed on the effects of BS on: i) the levels of ODA actually included in the budget process, ii) the composition of those funds and their predictability (see EQ 2.1), iii) the content and process of the policy dialogue, iv) the nature of technical assistance provision, and v) the composition and nature of development assistance flows as a whole (see EQ 2.2). The extent to which BS may have served to enhance harmonisation and to reduce the transaction costs of managing development assistance is the subject of EQ 2.3;
- an assessment of the combined (induced) outputs of BS and the corresponding government actions, in particular with regard to the changes that BS would normally be expected to induce in relation to the financing and institutional framework for public spending and the public policy process (EQs 3.1 – 3.3). Particular attention should be given to assess the changes (at national and/or sector level) which may or may not have been induced by BS as regards the quality of public policy making, macroeconomic management, the quality of PFM systems, allocative and operational efficiency of public expenditure, and the quantity and quality of public goods and services delivered. Evaluators should carefully assess the government's achievements, analyse the most important determining factors, and identify the specific BS contributions and related counterfactuals. The assessment must also take into account the non BS factors which influence government's outputs, including the stability and sensitivity of the political framework, the capacity of civil society organisations and their ability to interact with the government, the inherent capacities of the public sector, and the external economic environment. In highly decentralised countries attention should be given to policy changes (or their absence) at the local level. Outputs may be quite satisfactory at national level without required improvements at the provincial or local level.
- while the focus of the assessment is on the validity of the programme theory of budget support, attention should also be paid to possible unintended effects.

The findings of this step will typically rely on data from different sources, including:

- documentary review, encompassing: i) donors' strategic, policy and programming documents (country strategies, Memorandum of Understanding, financing agreements of the budget support programmes, disbursement & progress reports, technical assistance reports, etc.) ; ii) government development strategy documents, policy evaluations and other official documents; iii) documentation reflecting the process and content of policy dialogue and coordination with government and among donors; and iv) academic literature;
- compilation, reconstruction and analysis of datasets related to the BS programmes specified by type of agency, including: i) BS stated objectives (overall and specific objectives); ii) content of the policy dialogue and the procedures for dialogue; iii) planned and actual disbursements (amounts, timeliness of disbursements, procedures and criteria) by year and where possible by quarter; iv) disbursement conditions, performance indicators and related monitoring systems; and v) content of technical assistance and other capacity building support;
- compilation, reconstruction and analysis of datasets on: i) budgetary and fiscal trends ii) budgeted and actually realised revenues and expenditures broken down according to administrative, economic and functional classifications; iii) development of the PFM system; and iv) output trends of the sectors covered by the different BS programmes (e.g. macro- and micro-economic data and indicators);
- semi-structured interviews and in some cases focus group discussions with key resource people, among which are: i) donor representatives both at headquarters and in the field; ii) representatives of national

authorities and bodies (ministries, public agencies, National Audit Office, Public Accounts Committee and Audit Committee of the Parliament, etc.); and iii) non-state actors (civil society organisations, private sector, academic world).

Information and data collected from the various sources (documents, database, interviews, panel discussions, etc.) should then be cross-checked, i.e. analysed and compared (triangulation of data and sources of information) to verify the hypotheses and preliminary answers to the evaluation questions.

3.3. Step Two: assessment of outcomes and impact and their determinants

In Step Two, the evaluators will have to assess the economic and social outcomes and impact targeted by the government strategy and the BS programmes, and the factors which have caused and/or contributed to these outcomes and impact, both internal (i.e. linked to government policies, strategies and expenditures) and external. Step Two deliberately reverses the direction of the assessment from a direct causality analysis from level one to level three (= Step One) to an attribution analysis, tracing (backwards) the determinants of the observed outcomes and impact (= Step Two). Step Two will thus involve:

- a review of the impact and outcomes set as objectives by the government and supported and promoted by BS, so as to generate a summary of the key changes that have taken place during the period under evaluation. At the impact level the following issues are usually covered: i) economic growth, ii) the incidence of income and non-income poverty, iii) the degree of empowerment and social inclusion of disadvantaged groups, and iv) the degree of achievement of the sector specific development objectives as identified in the BS programmes²¹. At the outcome level the focus should be in particular on changes in: i) business confidence and competitiveness (private investment, trade, employment, etc.), ii) the appreciation and use made by the beneficiaries of the goods and services provided by the public sector (school enrolment, access to primary health care, etc.) and iii) the degree of confidence of the general public as demonstrated by internal security, tax payments and so forth.
- the identification of the factors that have contributed to determine the key changes (or absence of changes) at impact and outcome level²². The assessment needs to identify all major determining factors that in most cases will relate to government policies, strategies and expenditures, but it must also take into consideration factors wholly external to the government strategy, which might have had an influence on the outcomes and impact in either a positive or negative sense, such as the regional and international economic and political environment, and any unforeseen economic event that has had an impact on the observed changes²³. As regards government actions the assessment should seek to identify the positive linkages between these actions and improved outcomes, but it should also note where a failure to act or an inappropriate form of action might have jeopardised the outcomes.

The methodological approach of the evaluation for this step should be as simple and focused as possible, given the complexity of the themes involved and the need to focus the evaluation finally on BS effects (in Step Three). Standard policy analysis matrices may be made, using information from existing recent surveys and studies. Counterfactual analyses could be carried out to help assessing the determining factors of the

²¹ Additional areas to be considered may vary according to: i) the specific partnership framework and the related priorities established by the government and the main development partners, ii) the specific objectives of the BS programmes, and iii) specific issues highlighted in the ToRs.

²² Sometimes, the time frame of the evaluation and/or the availability of suitable data do not allow an appropriate assessment of the impact indicators. In such a case, the evaluation has to be focused on outcome indicators and building justified hypotheses on their sustainability and capacity to generate the expected impact.

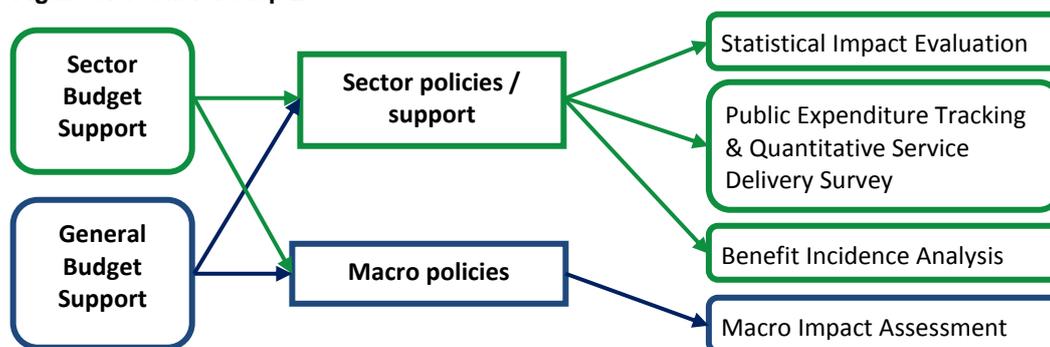
²³ Technically, it is not always necessary or possible to analyse all determinants. For instance, several techniques, like Randomised Control Trials allow to focus on the intervention itself. Nevertheless there may be two reasons why a broader analysis may be advisable. First, monitoring systems tend to assess the realisation of anticipated outcomes without paying attention to the contribution of the intervention. This may lead to an under- or overestimation of the impact of an intervention. Second, ignoring other determinants may lead to biased results if they are correlated with the intervention and with the outcomes and impact.

identified changes. The ToR of the evaluation should specify the extent to which counterfactual analyses will be required and feasible.

Typically, this step will need to be broken down into two main phases:

- a descriptive phase consisting of collection and review of quantitative and qualitative information and data as regards outcomes and impact of government policies promoted by the BS programmes. Data sets related to outcome and impact trends of the sectors covered by the BS programmes should be compiled and analysed (e.g. macro- and micro-economic data and indicators with specific focus on indicators related to the achievement of sector specific development objectives targeted by the BS programmes)²⁴;
- an analytical phase aimed at identifying / tracking the most influential internal and external factors that have contributed to the observed outcomes and impact of the government outputs targeted by budget support²⁵. There is no pre-defined set of tools to be used although this step should be as quantitative as possible. The selected mix of techniques will vary from evaluation to evaluation depending on the type and objectives of the BS programme to be evaluated, on the availability of data and on the budget available to the team for analytical activities²⁶.

Figure 5: Tools for Step 2



As indicated in figure 5, the main approach proposed for assessing the outcomes and impact of government sector policies supported by BS is statistical impact evaluation, completed with expenditure tracking and benefit incidence analyses. With regards to assessing the outcomes and impact of government policies supported by GBS or more generally the outcomes and impact at macro-economic level, it is suggested that different scenarios be developed with a view to identifying plausible impact of macro policies through a combination of quantitative and qualitative approaches²⁷. Examples of analytical tools used during the three pilot BS evaluations in Tunisia, Mali and Zambia are presented in Annex D. Ideally, as further discussed in chapter 5, these analyses and studies should be available at the start of the evaluation.

²⁴ The review of developments at the levels of targeted outcomes and impact will necessarily be limited by the quality of available data but in general should include an assessment of trends in relation to each of these variables, ideally disaggregated by gender, social group and geographical area and an analysis of the sector composition of growth.

²⁵ It should be stressed once more that the BS evaluation is not meant to be an overall assessment of either country performance or government strategies, but should be focused on both performances and strategies targeted by budget support.

²⁶ A balance needs to be struck between comprehensiveness and focus on key processes and developments. A practical approach would be to keep it fairly broad in first instance, then move to Step Three, and then if need be return to Step Two for further investigation of specific issues identified in Step Three.

²⁷ See also Compernelle, P. and de Kemp, A. (2009), Tools for 'Step 2': The evaluation of the impact of government strategies and the Paper 'Methodological details'.

3.4 Step Three: assessment of the contribution of BS to outcomes and impact

Finally in Step Three, the evaluators will have to synthesise and cross-check the results of the previous two steps. Step Three is therefore different from the other two steps and entails:

- the synthesis of Step One, which will allow the evaluators: i) to pinpoint the causal links between budget support and Government strategies and policy outputs, i.e. the effects of BS on institutional and policy changes, such as fiscal discipline, allocative efficiency in pro-poor expenditure, policy formulation processes, sector policy reforms, service delivery, etc. and ii) to highlight the roles played by the various BS components (funding, policy dialogue including disbursement conditions, and strengthening the institutional capacities through TA and/or other capacity building initiatives) and / or by other external factors;
- the synthesis of Step Two, which will allow the evaluators to highlight the actual determining factors of the development results supported / targeted by BS. Step Two will thus enable the identification of the different factors contributing to the achievements of the expected outcomes and impact set as objectives by the government and included in the BS programmes, including: i) the policy and institutional changes encompassing the effectiveness of the implementation of the policies, the effective involvement of local governments and implementing agencies, the operational efficiency of pro-poor expenditures, the fight against corruption, the association of civil society organisations in implementing government policies and their capacities, etc.; ii) those factors related to government action, which are not directly influenced by BS, e.g. the country political, economic and social dynamics and/or international partnerships which affect sectoral improvements; and iii) factors external to government actions, such as changes in international prices affecting rural incomes, regional economic and political dynamics or destabilisation (e.g. trade dynamics or crises in neighbouring countries), etc. The analysis therefore allows to establish possible causal links between government policies, strategies and spending activities and outcomes and impact targeted by the government.
- the comparison between the results of the previous two steps, which will allow evaluators to identify and discuss the 'transitive relation' between BS and the development results by highlighting consistencies, complementarities and possibilities of integration between the two steps and by assessing the significance of each of the chains of influence (BS → Government policies and interventions; and government policies and interventions → outcomes and impact targeted by BS). In this last step, the most significant BS effects (positive/negative/unexpected) highlighted in Step One will be compared with the most significant achievements (positive/ negative/unexpected) and the related determinants identified in Step Two. Possible strong or weak linkages will then be explored and the mechanisms through which they have (or have not) worked will be analysed;
- the verification and conformation (or rejection) of the hypotheses and conclusions formulated earlier by the team with reference to the above-mentioned chains of influence, through interviews with key resource persons and / or focus group discussions.

Recent experience gained during testing of the methodology has shown that relying on table formats focusing on the induced outputs level, facilitates presentation, cross-checking and comparison of the results of Steps One and Two and subsequently drawing conclusions in Step Three of the evaluation approach.

Recommended table formats are shown below (see tables 2, 3 and 4 hereunder and Annex D for an overview of the summary tables of the Tunisia and Mali budget support evaluations).

Table 2: Synthesis of the contributions of budget support and other internal and external factors to changes in government policies and strategies (Step 1)

Contribution of BS programmes	Government policies (induced outputs level in the CEF)	Other historical and/or policy interacting factors, and/or internal or external factors
Explanation of way(s) in which BS programmes have influenced / contributed to changes in government policy / strategy A, B, C, (etc.), including reference to the degree of influence and - where relevant - the type of influence (e.g. through the provision of funds, policy dialogue, strengthening of institutional capacities through the provision of TA and/or capacity building, or through a mix of BS components).	Government policy / strategy A, B, C (e.g. macro-economic stabilisation)	Explanation of way(s) in which other policy or internal/external factors X, Y, Z (etc.) have influenced / contributed to changes in government policy / strategy A, B, C, including reference to the degree of influence and - where relevant - the type of influence
Ranking of contributions (scale of influence): absent, weak, moderate, strong and very strong		

Table 3: Synthesis of factors determining changes observed at the outcome and impact levels (Step 2)

Government policies (induced output level in the CEF) having contributed to	Development Results	Other historical and/or policy interacting factors, and/or internal or external factors
Government policy / strategy A (e.g. macro-economic stabilisation) <i>including reference to the degree of influence</i>	Specific outcome/impact observed (e.g. macroeconomic stability and growth)	Other policy or internal/external factor X <i>including reference to the degree of influence</i>
Government policy / strategy B
Etc.
Ranking of contributions (scale of influence): absent, weak, moderate, strong and very strong		

Table 4: Synthesis of BS contributions to government policies (Step 1) and their contribution to development results (Step 2)

Type and intensity of the contribution of budget support to	Government policies (induced outputs in the CEF) having contributed to:	Development Results
Type and intensity of BS effects on government policy/strategy A	Type and intensity of change in government policy/strategy A (induced output)	Type and intensity of effects of (changes in) government policy/strategy A on the development results (outcomes and impact)
Type and intensity of BS effects on government policy/strategy B	Type and intensity of change in government. policy/strategy B (induced output)	Type and intensity of the effect of (the changes in) government policy/strategy B on the development results (outcomes and impact)
Etc.
Ranking of contributions (scale of influence): absent, weak, moderate, strong and very strong		

4 Organisational set-up of a budget support evaluation

4.1. Introduction

BS evaluations should be planned right from the start of the BS programme and should already be foreseen in the formulation documents and agreements. Pro-active planning of evaluations, timely collection of base line data, setting up of effective monitoring and information systems and identification of back-ground studies are all activities that can increase the quality of a BS evaluation, while reducing its duration and cost. These issues are elaborated in sections 4.2 and 4.3. of this chapter. Section 4.4 contains a number of observations on the scope and timing of a BS evaluation. As regards timing, the three steps of a BS evaluation could either be carried out in one limited (demarcated) period or separately at different times. The organisational set-up of a BS evaluation carried out in a limited period is elaborated in section 4.5.

4.2. Embedding the BS evaluation in the formulation and monitoring of the BS programme

In order to enhance the 'evaluability' of BS, the overall and specific objectives and the performance indicators of the BS programme(s) need to be clearly formulated, and possible differences between individual BS financing agreements need to be spelled out²⁸. Modifications of the set of indicators and targets, which may occur during the implementation of the programme, need to be well documented. Clearly formulated objectives, consistent Performance Assessment Frameworks (PAFs) and enhanced donor harmonisation as regards the design of the BS programmes and the PAF will make the reconstruction of the intervention logic of the BS programme(s) less time consuming. Moreover, in such a case the logic will be less complicated and thus easier to evaluate.

Secondly, it is recommended to strengthen the monitoring of the implementation of the BS programme(s) and of the Government's policies and strategies specifically supported by the BS programme(s), so that all data and information relevant for a BS evaluation are collected and stored during the implementation of the BS programme(s). This requires that a list is made of all crucial data to be collected during the implementation of the BS programme(s) before the BS programme starts, including references to when they should be collected (every month, every six months, every year) and by whom (institution of the partner government, non-governmental organisations, the donor chairing the BS group, etc.). If necessary, capacity building programmes aimed at strengthening the Monitoring and Evaluation Departments of the Ministry of Finance and Planning and of key sector ministries of the partner country should be launched in order to help setting up consistent databases and to ensure the availability and to improve the reliability of data needed to evaluate the PAFs and the BS programme at large. Such a capacity building programme could either be a specific component of the BS programme or be provided as complementary assistance in the form of project aid..

Furthermore, it is crucial that the policy dialogue and the monitoring of the PAF are well documented and archived throughout the implementation of the BS programme (annotated agendas, background documents, minutes of meetings, exchange of correspondence, etc.). This also applies to the BS harmonisation activities of the donors and the implementation of BS-related capacity development programmes.

²⁸ In case there is more than one BS donor, the term "BS programme" refers to the joint BS programme covering several Financing Agreements of different donors. This can be either a joint GBS programme or a joint SBS programme for a specific sector. The term "BS financing agreement" is used when reference is made to the contribution of a single donor to a joint BS programme.

4.3. The planning and preparation of BS evaluations

The timing and scope of the BS evaluation should be agreed and defined in a Memorandum of Understanding (MoU) of the joint BS programme, and an overview of data and information to be collected and background studies to be carried out during the implementation of the BS programme should be annexed to the MoU. The Evaluation Departments of the major BS donors and of the Ministry of Finance (MoF) of the partner country should be involved in drafting that specific section and the related annex of the MoU. The BS coordination secretariat - usually set up by the MoF – should be charged with preparing and facilitating the BS evaluation, coordinating the data collection, organising possible preparatory studies and surveys and liaising with other studies and surveys being conducted and of use for the BS evaluation.

The timing of the BS evaluation should be fine-tuned with (i) the implementation and finalisation of other evaluations, studies and surveys expected to provide important inputs for the BS evaluation and (ii) with other complex evaluations. In particular, the BS evaluation should be planned in such a way that maximum use can be made of the results of those other studies and surveys, while at the same time a congestion of (large) evaluations should be avoided. Some relevant evaluations, studies and surveys are:

- evaluations of the national growth and poverty reduction strategy;
- evaluations of the Country Assistance Strategy;
- evaluations of the Paris Declaration principles,
- sector development impact studies;
- household surveys;
- public expenditure reviews;
- public expenditure tracking surveys;
- assessment of the PFM system;
- etc.

In case it is envisaged to include specific sector case studies in the BS evaluation, it is important that the selection of the sectors takes place as soon as possible. Early identification of these sectors will allow focussing data collection efforts and possibly launching one or more special sector development impact studies to be carried out prior to the BS evaluation.

4.4. Defining the scope and timing of a BS evaluation

Keeping the BS evaluation approach as flexible and light as possible

Provided that monitoring and data collection are well organised and that supportive studies and surveys are being conducted during the implementation of the BS programmes(s) - as proposed above - most of the data and information needed for the Step One and Two assessments will be available at the time when the actual BS evaluation starts. This will facilitate BS evaluations and will make them less time consuming.

The definition of the scope (and level of detail) of the evaluation needs to take into account (i) the availability of data and information and (ii) the timeframe and budget of the evaluation. The choice as regards scope and level of detail should be reflected in the Comprehensive Evaluation Framework (CEF), while respecting the principle that the CEF should be focused on the main issues at stake.

Secondly – as already mentioned in section 3.1.2 - care should be taken to limit the number of key evaluation questions to a manageable number and to assure that each question is focused on a key issue (and does not represent a conglomerate of issues). Furthermore, each question should be accompanied by a set of SMART indicators and, if relevant, targets - which requires planning ahead for evaluation during the design phase.

Thirdly, it is recommended to be realistic with carrying out counterfactual analyses. There is a risk that those analyses become -in the case of BS evaluations - quite complicated and sophisticated, while in the end they may not produce conclusions which are more reliable than can be obtained through logical reasoning (because of unreliability of data and too many assumptions). It is advised to limit counterfactual analyses to cases for which sufficient reliable data are available and which do not need a series of assumptions.

Carrying out the steps of the BS evaluation in one limited period or at different times

The three steps of a BS evaluation could either be carried out in one limited (demarcated) period or separately, at different times with the obvious caveat that Steps One and Two must precede Step Three and that the three steps cover the same period of evaluation. Carrying out the steps at different times does not reduce the overall evaluation effort, but has the advantage of avoiding the concentration of all efforts in a limited period of time under one contract with the ensuing complex organisation and coordination processes (see section 4.5). On the other hand, separating the execution of the three steps has the disadvantage of not making use of a possible synergy and cross-fertilisation when Steps One and Two are carried simultaneously, while it also creates a risk of less coherence between the analyses of the three steps.

Separating the implementation of the three steps creates also the possibility of carrying out Step One more frequently than Steps Two and Three. Assuming that a comprehensive BS evaluation will take place only once in 5 to 7 years, Step One – which is a relatively light exercise, provided that an enhanced results-oriented monitoring system is functioning well – could be carried out every 2 years. This higher frequency will allow for providing timely information to the Partner Government and the BS donors on whether the implementation of the BS agreement is on track in terms of inputs, direct outputs and induced outputs and whether or not remedial actions are needed²⁹.

When the Step One assessment is carried out every two years and provided that most of the required data, information and a number of high quality supportive studies and surveys are available, the BS evaluation can then be focused on:

- verifying and validating the results of the Step One assessments;
- reviewing, validating and synthesising the findings and conclusions of supportive studies and surveys, which will be used as inputs for the Step Two analyses;
- consolidating the data and information of the monitoring systems needed for the Step Two analyses;
- carrying out the actual Step Two analyses by making use of – among others – the above mentioned inputs, and;
- conducting the Step Three assessment.

In order to provide a firm base for the Step Three assessment, it is advised that the results of Step Two are first discussed and validated before the Step Three assessment is carried out.

²⁹ It is to be noted that this would also provide information on outcomes and impact (insofar as these are part of the PAFs that trigger disbursements) but it would not allow to identify their determining factors.

4.5. Organisational set-up of a BS evaluation carried out in a limited period

4.5.1. Introduction

BS evaluations are joint donor operations. Thus donor coordination and the involvement of all significant donors in the evaluation process is of importance,

The organisational set-up of carrying out the entire BS evaluation (all three steps) in a limited period is summarised in the following subsections. It is assumed that one consultancy firm will be contracted for carrying out the entire evaluation. The phases and activities of such an evaluation are described in sub-section 4.5.2 and the key deliverables, the expected duration of the evaluation and the staffing of the evaluation team are presented in sub-sections 4.5.3 to 4.5.5. The indicative calendar of such an evaluation is summarised in table 5.

Should the three steps of the BS evaluation be carried out separately at different times (see section 4.4) – and possibly by different entities – the evaluation process will be different. In that case, the content of the following sections could be used as a reference and starting point for the definition of an alternative planning and phasing of the different steps.

4.5.2. The phases and activities of a budget support evaluation

When the BS evaluation is carried out as one single assignment to be carried out in a limited period, the work can be divided into six phases: (i) preparation, (ii) inception, (iii) data collection and preliminary analyses, (iv) in-depth analyses and validation, (v) synthesis and (vi) dissemination. The details of each of these phases are outlined in the following paragraphs.

The preparatory phase

The objectives of the preparatory phase are:

- to assess and seek to ensure the basic readiness for the evaluation (availability and quality of existing data, info on policy dialogue, other important projects to be included, relevant capacity building operations and priority sectors)³⁰ ;
- to ensure the active participation of the budget support donors, the partner country Government and representatives of key institutions in the partner country (e.g. Central Bank, research institutions, non-state actors ...);
- to clarify the roles of involved parties, specifically donors and the Government;
- to draft the terms of reference of the evaluation, including the definition of the temporal, geographic and thematic scope of the evaluation³¹ as well as its time frame and key sectors for in-depth analysis;
- to define and set up the management structure for the evaluation that suits the evaluation context and provides for a manageable process. This will typically comprise a Management Group and a Reference Group. The Management Group consists of representatives of the donor(s) providing budget support

³⁰ It should be noted that the review of developments at the level of targeted outcomes and impact will necessarily be limited by the quality of available data. To this end, coordinating the scheduling of the evaluation with the availability of periodical surveys on poverty and growth should be envisaged where possible. It should also be noted that during the preparatory phase, the check of the availability and quality of data will still be of a preliminary nature. A more comprehensive overview will only be available at the end of the Inception Phase.

³¹ Should the BS programmes cover various sectors, donors and the Partner Government will have to discuss the possibility of identifying a more limited set of key sectors for in-depth analysis / sector case studies.

and of the partner country's Government and is usually chaired by one of the lead donors. It is charged with managing the evaluation process effectively. The Reference Group will be constituted in the partner country and will consist of representatives of the partner Government and of the donor offices in the partner country. The group will usually be co-chaired by one of the Government representatives in the management and the chair of the Management Group. The Reference Group could include also other stakeholders (e.g. civil society groups or control bodies) in order to strengthen the transparency and the dissemination of the results of the evaluation. The group is used to provide expertise.

- to launch – once the above requirements are fulfilled - a tender for the selection of a consultant to carry out the evaluation and to select the winning proposal. Specific procurement processes will depend on the rules of the commissioning agency(ies)

The preparatory phase will come to an end when the contract with the consultant for carrying out the evaluation has been signed. Usually the main contract is financed by a single donor. Where feasible, co-financing is, however, the preferred option, and it can be useful for ensuring buy-in and commitment of different stakeholders.

The inception phase

The inception phase – the first phase of the actual evaluation - is aimed at structuring the evaluation and will entail:

- i. the collection of documentation and a preliminary desk-based review of those documents;
- ii. an inception mission aimed at:
 - identifying further data sources, collecting available documentation and data and making arrangements for the compilation / preparation of data in areas where there appear to be gaps³²;
 - discussing with the Management Group key features of the evaluation, expectations and if necessary clarifying the management arrangements of the evaluation;
 - establishing contacts with the relevant ministries and institutions in the partner country, and;
 - holding a workshop to present and explain the Three Step Approach to country level stakeholders (provided this is foreseen in the ToR).
- iii. an inventory encompassing BS programmes and other relevant / related programmes, including the compilation of datasets related to the inputs of BS programmes;
- iv. a general context analysis, including an overview of the country political economy in a historical perspective, the country's international position, and basic country level data (economic, financial and social);
- v. identification of the main specific features to be covered by the Comprehensive Evaluation Framework, and the ensuing drafting of the Intervention Logic(s) on which the BS programmes are (were) based;
- vi. selection of a limited number of sectors to be addressed in detail during the evaluation (if not yet done in the ToR)³³;
- vii. definition of a framework of (i) about ten to twelve EQs adapted to the Intervention Logic of the BS programmes and embracing the selected sectors (ii) possibly sub questions accompanying the EQ, and (iii) indicators to be used for answering the questions;
- viii. drafting the methodological design, including evaluation tools to be applied in Steps 1 and 2 and the related sources of information together with indications of any limitations / risks envisaged in the retrieval of data and application of those methods as well as corresponding mitigation strategies;
- ix. drafting a detailed work plan for the next phases of the evaluation.

³² The use of a questionnaire aimed at collecting preliminary information and documentation on GBS/SBS programmes conducted by each BS donor in the country should be assessed at this stage. Information to be gathered would centre on the different aspects of GBS/SBS programme to be included in the inventory.

³³ If the sectors are already identified in the joint BS programme(s) and/or the financing agreements of the individual donors, and as such mentioned in the ToRs, focus will be placed on drafting the sectoral Intervention Logic.

Towards the end of the inception phase, the Evaluation Team will submit an **inception report**, which will include the above mentioned outputs. The report will be discussed with - and reviewed by - the Management Group. The inception phase will come to an end with the approval of the inception report by the Management Group.

Further data collection and preliminary analyses

This second phase of the evaluation consists of further data collection, carrying out preliminary analyses and formulating preliminary answers to the EQs and the related hypotheses (which will be tested during the third phase of the evaluation)³⁴. In order to do so, the evaluators will need:

- to examine the validity of the Intervention Logic(s) of the BS programme(s) being evaluated (Step One) with a view to provide the preliminary answers to EQs at levels one, two and three (see section 3.2 and annexes A and B for further details on tools and sources of information), and;
- to identify and assess the economic and social outcomes and impact targeted by the government and BS programmes, and identify the main determining factors (government policies, strategies and expenditures and external factors) (Step Two) with a view to provide the preliminary answers to EQs at levels four and five (see section 3.3 for further details).

This phase will thus consist of:

- a more detailed desk-based review of documentation;
- the compilation, reconstruction and analysis of datasets;
- a first set of semi-structured and open-ended interviews with key resource persons;
- a context analysis (including a political economy analysis in a historical perspective) in the selected sectors to complement the analysis undertaken during the inception phase;
- an assessment of determinants of macro-economic changes;
- an assessment of changes at the micro-level (for different sectors).

This preliminary analyses phase will also include a field mission in order: i) to collect additional data, ii) to carry out interviews and discussions as regards the various analyses with country based stakeholders; and iii) to discuss preliminary findings with the Reference Group.

Towards the end of this second phase, the evaluation team will write an **interim report**, which will include:

- updated (or simply confirmed) sets of outputs already mentioned in the inception report, namely: i) the inventory of the BS programmes and other relevant programmes; ii) the general context analysis, which will be integrated through the context analysis in the priority sectors; iii) the Intervention Logic(s); iv) the final set of evaluation questions (and sub questions), including preliminary indicators and, where relevant, preliminary qualitative and quantitative targets; and v) the methodological design, including the tools and methods to be used;
- the evidence from the detailed assessments and analyses carried out for both Steps 1 and 2;
- the preliminary responses to the EQs and the hypotheses to be tested later on, and;
- a plan for data collection in the partner country (both in the capital city and at decentralised level), specifying the sources, indicating any limitations and describing how the data will be cross-checked.

This phase will end with a meeting of the Evaluation Team and the Management Group where the interim report will be discussed. A final version of the interim report and formal approval of such a report are not needed because of the preliminary status of its content. The draft of the interim report and the comments received are just inputs for the next phase.

³⁴ This third phase has sometimes been called the "desk phase", but in none of the three pilot evaluation it actually consisted of only desk work.

In-depth analyses and verification.

This third phase – which is largely carried out in the partner country and therefore traditionally termed the field phase - is devoted to the collection of additional data needed to fill in gaps, further in-depth analyses of the various issues and the verification of the preliminary responses and hypotheses formulated in relation to each EQ during the previous phase. At this stage of the evaluation, most of the underlying analyses of available data for the evaluation will be completed and all remaining data gaps will be addressed. Specific tools will be used to collect missing information. Applied tools for data collection and analysis will include (as appropriate) interviews with key stakeholders and related institutions, focus groups, surveys, expert panel, case studies, multi-criteria analysis, etc. (see annex B for more details). The output of the use of these tools will be combined to ensure rigorous conclusions and operational recommendations.

A large part of the evaluation work, especially the work concerning BS inputs and their effects (direct and induced outputs) on public policy processes and management, is likely to be carried out in the capital city of the partner country. However, in order to capture different perspectives and to substantiate findings by triangulation, the evaluation team will also visit and interview staff of local governments and representatives of the private sector and civil society organisations at decentralised level. Furthermore, it will be of utmost importance to capture the perspective of citizens as the users of public services and the beneficiaries of public policies. The opportunity to obtain a genuine ‘field perspective’ in this respect is essential, especially for gaining knowledge about the outcomes and impact (levels 4 and 5) of the Government policies and spending. It is also important to approach the same evaluation question from different angles and perspectives and to try getting confirmation of findings from different sources (central government, local governments, donors, citizens, private sector, etc.)

This third phase will be concluded by meetings with the Reference Group in the partner country and with the Management Group, where the findings and preliminary conclusions will be presented (in the form of an **aide memoire** and a **Power Point presentation**) and discussed. These presentations and discussions will focus on:

- the verification process (including problems encountered and solutions adopted, coverage and reliability of data), highlighting the key points discussed and the amendments suggested to the responses to the EQs;
- the updated evidence resulting from the detailed assessments and analyses carried out for both Steps 1 and 2, and;
- a preliminary synthesis of the results of Step 1 and Step 2 (table formats as outlined in section 3.4), and the ensuing preliminary identification of transitive relations (or lack thereof) between budget support and the development results (Step 3).

The synthesis phase

Following the de-briefing meeting with the Reference Group and the Management Group, the evaluation team will start working on the overall analysis of the collected information. The synthesis phase is focused on completing Step Three of the evaluation approach and can usually be carried out in the form of desk work at the place of residence of the consultants. Due to the complexity of a budget support evaluation, often including the evaluation of GBS and SBS, an overall view and thorough analysis of all the information collected are essential, be it information based on documents, data, facts and figures from interviews, case studies or outcome and impact studies on certain thematic aspects. During this phase the evaluation team will bring together in a comprehensive manner all the collected information, findings and preliminary conclusions formulated throughout the evaluation process. Furthermore, the evaluation team will also formulate final and overall conclusions (particularly related to Step 3), identify lessons learned and formulate recommendations.

Based on the overall analysis and synthesis of information, the evaluation team shall prepare a first draft of the evaluation report which fulfils the objectives of the evaluation and is structured according to what is prescribed in the terms of reference of the evaluation³⁵. This first draft will then be sent to the Management Group. Usually the budget of the evaluation does not allow for another visit of the team leader to the partner country to discuss the draft report with the Reference Group. In that case the group will be asked to provide written comments. After receipt of the comments from the Reference Group and the Management Group, the Evaluation Team will meet and discuss the draft report with the Management Group. Subsequently the Evaluation Team will produce the **final report**, taking into account the comments received.³⁶

The dissemination phase

Finally, it is essential that the final report is presented at headquarter level as well as at a **national seminar** in the partner country to the Government, the donors (development agency officials), key national stakeholders involved in budget support and to the wider community of political leaders, government officials, academics, civil society organisations, private sector representatives, and others to whom the findings and recommendations could be of interest. The national seminar may be held before the final meeting of the Management Group and the adoption of the final report by the contracting donor.

4.5.3 Key deliverables

The key products to be delivered by the consultant are:

- the inception report (phase 1);
- the interim report (phase 2);
- the aide memoire and the Power Point presentation (phase 3);
- the final report (phase 4), and;
- the minutes, conclusions and recommendations of the national seminar (phase 5).

4.5.4. Duration

The duration of an evaluation will depend on its scope, management (number of partners involved) and resources available. It is of importance of being realistic regarding the timeframes and on managing expectations of various stakeholders. There may be strong pressure to produce quick results, so that those involved in managing the evaluation need to communicate clearly about the process and expected results in order to help managing the expectations. Accordingly, if sufficient resources are not available, a BS evaluation should not be carried out at that time.

It can be expected that a budget support evaluation will take normally about 12 to 18 months. In case most of the needed data and documentation are available at the beginning of the evaluation, the duration may well be limited to 12 months. An evaluation lasting longer than 18 months is losing its value, because the longer the gap between the period of evaluation and publication of the report, the lower the relevance of its conclusions and recommendations for actual policy making and implementation of on-going and future BS programmes.

³⁵ Alternatively two separate reports could be made: a main report and a report with the detailed findings of the Steps 1 and 2 analyses.

³⁶ The evaluation team may either accept or reject suggestions and recommendations formulated by the Reference Group and Management Group to modify the text. The team should specify what changes were made in response to comments and should explain why certain suggestions and recommendations were rejected. The comments received and the responses from the Evaluation Team will be sent to the Reference Group and the Management Group together with the final version of the report.

Given the time necessary for preparation and contracting, planning of the evaluation should begin approximately 2 years before it is supposed to be finished. An indicative calendar is provided in table 5 (see next page).

When planning an evaluation, timing of the conclusions and recommendations should consider their relevance for actual policy making and implementation of on-going and the design of future BS programmes.

For instance, the planning should ensure that an evaluation will be completed ahead of key policy reforms or renewal of partnership agreements. Often preliminary findings and research can be shared throughout the process (rather than waiting for the final report), which can help in policy formulation and keep stakeholders engaged and interested.

4.5.5. Staffing of the evaluation team

The level of resources required to carry out a BS evaluation depends on the scope, scale and number of GBS and/or SBS programmes to be evaluated. Yet, even with relatively simple evaluations, it is necessary to assemble a team which combines: i) knowledge of evaluation techniques including statistical analyses; ii) knowledge of the country in question; iii) expertise in macroeconomics, public finance management and political science; and iv) knowledge of the specific sectors to be analysed. In most cases a team of at least five to six people is needed preferable both national and international consultants/ researchers of which at least three seniors. In total about 350-450 person days will be required for accomplishing the evaluation. For very complicated evaluations covering a very diversified set of sectors, more team members and more person days may be needed.

Table 5. Indicative calendar of a budget support evaluation carried out in a limited period

Evaluation phases and stages	Notes and reports	Months	Meetings/Communications
Preparation phase			
Assessment of readiness for the evaluation	Notes		
Establishment of the Management Group and the Reference Group	Notes		
Drafting of the ToR	Draft ToR		
Approval of the ToR	Final ToR		
Procurement and award of contract	Technical and financial offer		Formal contract between the consultant and the contracting donor agency
1. Inception phase			
Kick off meeting			Meeting with Management Group
Preliminary desk review		<i>N</i>	
Visit of Team Leader to the partner country. Interviews with key stakeholders			Meeting with Reference Group in partner country.
Preparation of inception report	Draft inception report	<i>N</i>	
Review of inception report		<i>N+1</i>	Meeting with Management Group
Drafting of final inception report	Final inception report	<i>N+2</i>	Approval of inception report by Management Group
2. Data collection and preliminary analyses			
Detailed desk review		<i>N+2</i>	
Visit of Core Evaluation Team to partner country. Interviews with stakeholders		<i>N+3</i>	Meeting with Reference Group in partner country.
Preparation of interim report	Draft interim report	<i>N+4</i>	
Discussion of interim report		<i>N+5</i>	Meeting with Management Group
3. In depth analyses and verification			
Visit of evaluation team to partner country: - further data collection, in depth interviews and panel discussions - presentation of preliminary findings	Aide memoire & Power point presentation	<i>N+6-7</i> <i>N +7</i>	Meeting with Reference Group in <i>partner country</i> Meeting with Management Group
4. Synthesis phase			
Writing draft report	Draft report	<i>N +8-9</i>	
Review of draft report			Meeting with Management Group
Receipt of comments	Notes with comments from Management and Reference Groups	<i>N +9</i>	Comments consolidated and sent by the Management Group
Writing final report**	Final report	<i>N +10-11</i>	Approval of final report by Management Group
5. Dissemination phase			
Preparation of the seminar	Power point presentation	<i>N+12</i>	
National seminar (s)	Power point presentation and discussion.	<i>N + 13</i>	One day conference. Participants: Reference Group, Management Group, Headquarter Evaluation Units + donor representatives, various stakeholders in partner country.

Annex A

Tools for Step Two of the Budget Support Evaluation

This annex provides a brief overview of the different types of analysis carried out by the three pilot evaluations distinguishing where necessary and relevant, between the different outcome and impact levels³⁷. It is reminded that the choice of techniques used in each evaluation was mainly guided by the availability of data and by the level of funding available to each team for analytical activities. Types of analyses undertaken during Step 2 included:

- Political economy analysis in a historical perspective encompassing economic literature review, qualitative data analysis and interviews with high level resource persons. In Mali and Zambia it was particularly useful to assess sector reform processes. In the Zambia evaluation, for instance, this analysis was used for an assessment of the conditions for effectiveness of BS at the sector level. In Tunisia, apart from its use in Step 1, it helped identifying a number of historical factors that contributed to achievement of the development results and interacted with the government policies targeted by BS.
- An assessment of determinants of macro-economic changes:
For the Tunisia evaluation several quantitative techniques have been used for an assessment of the impact of budget support on economic growth in Tunisia, including:
 - dynamic panel data analysis, including five countries of the region which had signed Association Agreements with the EU. This analysis helped to improve the understanding of the relationships between specific policies over time and the *intended* results in terms of economic development;
 - cluster analysis, covering six countries of the region which had signed Association Agreements with the EU. This analysis was used to identify the specific factors linked to economic and social growth in Tunisia as compared with neighbouring countries;
 - fixed effects regression to identify the relation between two dependent variables (GNP *per capita* growth and the Human Development Index) and a number of country macro-economic and sector variables related to policy and non-policy factors.

The Zambia evaluation simulated the macroeconomic effects of aid inflows in Zambia using a General Equilibrium model that was developed by Clausen / Schürenberg (2009).

³⁷ These are extracts from the “Joint Methodological Note”

▪ An assessment of changes at the micro-level (for different sectors):

The three evaluations applied different quantitative techniques for an assessment of the impact of interventions (investments) in the sectors of interest:

- Contribution analysis, based on a literature review, analysis of monitoring reports, interviews, focus group discussions (with private sector representatives, civil society organisations and intended beneficiaries) and a qualitative analysis of survey results. In the case of Mali it is worth mentioning that, in addition to relying on existing surveys, separate funding was made available to undertake a 'Local Level Perception Survey'.³⁸
- Correlation analysis between the main outputs targeted by the national health policy and the main service delivery outcomes (Mali).
- Regression techniques:
 - Difference-in-difference analysis to assess the relative importance of the different outputs of government spending in the education sector in Mali, and to measure the impact on enrolment of investments in the education sector in Zambia.
 - Fixed effects regression analysis undertaken in Mali as part of the education case study to reveal the major determinants of two key proxy variables for quality: (i) the repeater rate (within primary) and (ii) the primary-secondary transition rate. In Zambia it was used in the Health, Education, Water and Sanitation and Roads sector case studies. The Zambia evaluation of the education sector included a number of regression techniques simultaneously: random effects regressions (in order to exploit the full richness of the database, fixed effects regressions (focusing on changes over time) and ward fixed effects.
- Propensity score matching (PSM) was used in the agricultural, roads and water & sanitation case studies in Zambia. For the water and sanitation and roads cases studies, PSM was mainly used as a kind of a robustness check of the regression results. The agricultural sector case study combined PSM with a difference-in-difference analysis, which allowed controlling for time-invariant effects.
- Benefit incidence analysis: the Zambia evaluation used this technique for an assessment of the beneficiaries, the allocation of resources and of improved service delivery. The analyses were carried out at household level, district level (focusing on district poverty rates) and urbanisation level (urban/rural differences).

³⁸ The survey covered 19 communes in 4 different regions and allowed to shed light on important questions, which available quantitative data did not allow to answer (e.g. underlying causes of given results highlighted through econometric analyses such as 'was the quality of teaching suffering as a result of fast expansion?' or 'why teachers on short-term contracts achieved lower repeater rates and higher transition rates into secondary schooling than regular fully trained teachers'.

Annex B: Tools for the verification phase

Specific tools could be used to validate preliminary answers and hypotheses in relation to each EQ and to collect missing information. Some of the tools that could be used are:

- Semi-structured individual interviews based on clear guidelines and check lists. Stakeholders interviewed could include: key officials of central ministries and agencies (Finance, Planning, Office of the President, etc); relevant sector ministry staff at national, regional and district levels; representatives of the development partners involved in budget support, members of the national parliament; oversight bodies, Central Bank staff,; academics; and private sector and civil society resource persons.
- Group interviews / focus groups. Organisation of one or more thematic meetings or workshops with representatives of the different categories of stakeholders in order to collect the same type of information from different persons at the same time, to increase the validity and the overall understanding of certain issues. Focus groups with members of private sector associations and other relevant sectors of civil society based on clear guidelines and check lists would be useful as well.
- Case studies/outcome and impact studies. Undertaking of specific case studies or outcome/impact studies on sectors/themes identified when providing preliminary answers to the evaluation questions, to complement the broader partner country analysis.
- Workshops aimed at sharing preliminary results of data analysis with relevant stakeholders so as to ensure completeness and correctness. These workshops can prove to be particularly helpful in cross-checking the team's understanding of events, in resolving apparent inconsistencies, and in identifying additional relevant data and reports.

Annex C: Example set of key Evaluation Questions (EQs)

Step	Level	Key evaluation questions	
1	1	1.1. To what extent was the design of the BS programme(s) appropriate and relevant in view of the political, economic and social context in the country, the government's policy framework and the external Development Partners development assistance strategies?	
	2	2.1. To what extent has budget support contributed to an increased size and share of external funding subject to the government's budgetary process and to an improved predictability of aid flows?	
		2.2. To what extent has budget support contributed to an efficient and effective policy dialogue and to well-coordinated capacity building activities, both focused on strategic government priorities?	
		2.3. To what extent has budget support contributed to greater harmonisation and alignment and reduced transaction costs of external assistance?	
	3	3.1. Have macroeconomic management and PFM been improved, and to what extent and through which mechanisms (flow of funds, policy and institutional effects, others) did BS contribute to these improvements?	
		3.2. Have there been improvements in policy formulation and implementation processes, in the quality of the policies and in democratic accountability, and to what extent and through which mechanisms (flow of funds, policy and institutional effects, others) has BS contributed to these improvements? <i>(This EQ will have to be adapted and focused on the specific policy areas targeted by the BS programmes).</i>	
		3.3. How has the level and composition of public spending changed, and with which consequences in terms of allocative and operational efficiency in the provision of public services and goods, and what was the contribution of BS to the observed changes?	
	2	4 + 5	4.1. To what extent have there been changes in sustainable and inclusive economic growth and the structure of the economy as targeted by government and supported by BS? To what extent are these changes due to economic policies, structural reforms and changes in allocative and operational efficiency of public spending and other external and internal factors?. <i>(In case only SBS has been provided, the EQ should be focused on sector performance and the relationships between the overall economic context and sector performance).</i>
			4.2. To what extent has there been an increase in the use of public services and what has been the impact on the overall livelihoods of the population, in particular as regards health, education, water and sanitation, and to what extent can the increased use of public services and improvements in overall livelihoods be related to changes to government policies or policy processes and/or to other external or internal factors? <i>(For SBS, this question should be focused on the sectors concerned).</i>
			4.3. To what extent have there been changes in income and non-income poverty, and to what extent can these changes be related to changes in economic and social policies, structural and institutional reforms and public spending and/ or to other external or internal factors?
4.4. To what extent can significant changes be identified in terms of (i) improved confidence of the population in the performance of the Government particularly as regards the budget process and service delivery and (ii) other objectives agreed in the BS agreements (e.g. improvements in environmental protection, democracy, rule of law, human rights and other governance issues), and to what extent can these changes be related to changes in government policies and policy processes and/or to other external or internal factors?			

3	All	<p>5. To what extent – and through which mechanisms - has BS contributed to the attainment of the development results identified in Step Two?</p> <p>This question has to be answered by comparing the results of Step 2 and Step 1. More specifically by looking at the coherence and complementarity or discrepancies between the results of the two steps</p>
---	-----	--

Annex D: Summary tables of the Tunisia and Mali budget support evaluations

Overview of summary tables prepared for the Tunisia and Mali pilot BS evaluations.³⁹ Although Step 3 has been carried out in slightly different ways (including different processes, table formats and different scales to assess influence), in all cases it has allowed summarising, synthesising and double-checking the results of the evaluations. In Zambia, no summary table was produced as Step 3 was carried out as part of a continuous exchange between members of the evaluation team, later consolidated during a two-day workshop when the team discussed Step 1 and sector reports together with external experts.

Table C.1: Summary of Budget Support contribution to Government of Tunisia policies (Step 1) and their contribution to development results (Step 2)

Intensity and type of Budget Supp. contribution* to	Govt of Tunisia Policies having contributed to..	Development Results	Other historical and/or policy interacting factors	
Dialogue/TA/Funds (Moderate to Strong)	Macroeconomic stabilisation	Macroeconomic stability, growth and poverty reduction	Long experience of Structural Adjustment with WB/IMF	Opportunity framework provided by the EU-Tunisia Association Agreement
Dialogue/TA (Strong)	Trade liberalisation	Trade and economic growth and poverty reduction	Experience and propensity toward export-led models	
Dialogue/TA/Condit. (Very Strong)	Private sector development and restructuring of financial sector	Moderate investment & employment growth, enhanced competitiveness and incomes	Resistance of national monopolies in domestic market	
Funds/TA (Strong)	Upgrading of education system	High increase of access & equity, moderate increase of quality & employability	Vocation toward social modernisation, but lack of SWAP	
Dialogue/TA/Funds (Strong)	Improving PFM	Macroeconomic stability and growth	Long experience of Structural Adjustment with WB/IMF	
Dialogue (Moderate)	Improving public expenditure transparency and accountability	Good information, but limited participation of citizens	Very centralised state and controlled participation	
Dialogue/Condition. (Moderate)	Improving other cross cutting policies (gender, justice, environment)	Low increase of domestic competition	Resistance of national monopolies in domestic market. Other historical factors	
*Scale of influence in five levels described as follows: absent, moderate, moderate to strong, strong and very strong				

³⁹ The tables presented here are summary tables prepared for the Methodological Note (December 2011). The specific country reports present a number of tables (by evaluation step and/or by sector) which provide a synthetic base for Step 3 crosschecking and conclusions.

Table C.2: MALI – Government-related determining factors of the socio-economic outcomes and impact and their link with Budget Support

Budget Support inputs with a potential influence on the factors determining final outcomes and impact	Degree of Influence of budget Support input	Factors determining the changes observed at the Outcome/ Impact level	Degree of Influence of the determining factor	Specific Outcome/ Impact in question
NB. External factors excluded. Hence, analysis limited to the Government Interventions identified amongst the Determining factors in Step Two. (Investments/ specific types of recurrent expenditures, policy decisions, institutional reforms, etc.)				
GBS dialogue and conditions on Macroeconomic policy	Negligible (in relation to the established government commitment to the WAEMU convergence criteria and the targets agreed with the IMF)	Prudent macroeconomic management by the Authorities, notably regarding control of the fiscal deficit and the inflation rate.	Important	5% per annum average rate of real growth in GDP over evaluation period (the highest of the WAEMU zone within the period)
Budget Support finance (SBS & GBS) equivalent to up to 13.6 % of annual public spending, disbursed with an acceptable degree of predictability	Important	Reduction in the fiscal deficit after grants to an average of 0% of GDP over 2003 to 2008 (against an average of - 1.6% over 2000-2002)	Important (in particular because of its moderating influence on monetary growth and the real interest rate)	
GBS dialogue on the PRSP (CSLP/ CSCRP) and related strengthening of capacities for PRSP implementation	Moderately important (GBS dialogue has been important in creating a robust annual monitoring framework for the PRSP but Government commitment to the funding of the PRSP priority sectors was already well established before GBS)	Improvement in the allocation of resources to the PRSP priority sectors	Weak (The composition of growth does not suggest that Government sector policies and programmes were determining factors in the rate of economic growth achieved)	
Budget Support finance (SBS & GBS) equivalent to up to 13.6 % of annual public spending, disbursed with an acceptable degree of predictability.	Important	Increase in the level of own financed Public Investment from 17% of Budget revenues in 2002 to 21 % in 2009, peaking at 28% in 2007	Moderately Important (as a counterweight to the continued stagnation of private sector investment)	
GBS dialogue and conditions on PFM reform and related stimulus to PFM capacity building	Weak (PFM reforms are not sufficiently advanced to have generated significant efficiency effects)	Improved efficiency in Budget execution, in particular for Public Investments	Absent (Faster and more efficient execution of public spending would in principle enhance the impact of public spending but there is no evidence of such efficiency effects)	
GBS dialogue and conditions on the «Doing Business» indicators and related aspects of the private sector investment	Weak (Dialogue has not focused on obstacles to investment identified by the domestic private sector)	Improvement of the Investment Climate	Negligible (The rate of private investment has remained low and the chief constraints on investment commonly identified remain present)	
Scale of influence in five levels, described as follows: absent, negligible, weak, moderately important, and important.				



OECD Development Assistance Committee

Network on Development Evaluation

2, rue André Pascal

75775 Paris Cedex 16

Paris, France

www.oecd.org/dac/evaluation