

# Evaluating General Budget Support Issues Emerging from Evaluation Findings

## *Workshop Report*

### **Background and purpose**

There is growing interest in General Budget Support (GBS) as a modality of development cooperation. In a number of developing countries external assistance is increasingly taking the form of GBS, with a relative decline in both project aid and sector- or area-based programme support. The Evaluation community is therefore asking when and how it will be most appropriate to commission evaluation work to assess the viability and effectiveness of this type of aid.

This workshop, hosted by DFID on behalf of the DAC Evaluation Working Party, brought together evaluation specialists and other representatives from a wide range of bilateral and multilateral agencies to undertake a preliminary discussion of this question. The workshop reviewed the findings and lessons from a number of recent studies concerned with different aspects of the evaluation of GBS. Ideas and proposals were exchanged about how best to move forward towards an agenda of joint evaluation and research on GBS.

The workshop agenda included five formal presentations, a panel discussion and group work. It was introduced by Barrie Ireton, DFID Director General, and chaired by Simon Maxwell, Director of the Overseas Development Institute, London. The presentations were as follows:

- Mick Foster, *Aid effectiveness and the choice of aid instruments*
- Howard White, *Evaluating programme aid: some lessons from experience*
- Tony Killick and Charles Harvey, *Independent evaluation of SECO's budget support: lessons learnt*
- Andrew Lawson and David Booth, *General budget support evaluability study for DFID*
- Geske Dijkstra, *Methodology for evaluating Dutch debt relief via country studies*
- Paul Balogun, *What we have learned from the joint evaluation of the Comprehensive Development Framework*

The following report identifies some highlights from the presentations in the boxes. The main text contains a summary of the discussions that took place around these contributions and the workshop theme – how to move forward towards joint evaluation of general budget support.

## GBS and programme aid: what have we learned?

A major theme of the opening sessions was: what is new and what is not new in the field of budget support? Programme aid has been a variable but always important component of aid flows since the 1960s. But there is a family of programme aid instruments, and there are new forms of it, which need to be the focus of discussion today.

Some important observations apply to all recent forms of programme aid (e.g. the resource transfer takes place when the forex is sold; what matters after that is the *relationships*, not exactly what is done with the money). Other issues about GBS are specific, and arise from the particular thinking about how best to support countries with both liberalised foreign exchanges and PRSPs.

### Box 1: Barrie Ireton on cycles in UK aid and budget support

Budget support is both old and new – variants include:

- “budgetary aid” (immediate post-colonial);
- “commodity aid” (1970s);
- balance of payments support – pre- and post-liberalisation variants;
- SWAps;
- GBS, focused on whole budget, supporting PRSPs and MTEFs.

Long-term cycle of motivation: crisis response → development concerns → crisis response → development concerns

### Box 2: Mick Foster on GBS and current aid issues

Characteristics of GBS modality:

- conditionality concerns macro *and* budget;
- earmarking nominal or none;
- accountability through recipient government systems.

When to use it? Main choices can be set out as a decision-tree (see CAPE Working Paper 158).

GBS in the current aid context (after Monterrey commitments, and the aid effectiveness/selectivity debate):

- Bigger aid volumes + more selectivity → diminishing returns + fewer poor people served?
- Are donors causing poor policy environments? (by amplifying shocks, by-passing systems)
- High aid can → growth and reduced dependence, where countries have strong, centralised PE planning and budget systems (Taiwan, Botswana, Uganda)

Evaluation issues:

- Need for country case studies to unpick complex relations between aid, policy, selectivity, effectiveness
- Focus on strengthening the institutional environment, rather than current policies?
- What works to reduce unpredictability?

Following the initial discussions, many participants considered some issues and principles to be well established. For example:

- budget support should be articulated as closely as possible with the budget cycle of the recipient, to promote rational decision-making and democratic accountability;
- as little aid as possible should be off-budget;
- improving predictability is a

crucial issue – this may mean having clearer guidelines about where to start GBS and the initial risk assessment; specifying up-front how often and under what conditions disbursement will be restricted; looking to the long term and making a big effort not to be blown off course by short-run storms; being realistic about

political constraints at both ends of the relationship; defining predictability not as an invariant flow but as agreed rules; not going too far down the road of outcome conditionality; focusing on the level that matters most – predictable disbursement to line ministries and local governments from the centre; and having contingency plans – what to do when things go wrong;

- delivering GBS should including providing demand-led technical assistance – not just writing a cheque;
- joint design, monitoring and evaluation is crucial, and should use country systems as far as possible; cooperation and harmonisation between bilaterals and IFIs is particularly important, to reduce transaction costs to recipient;
- Tanzania’s Independent Monitoring Group provides a variable ingredient of reciprocal monitoring – a good way of tracking the quality of aid delivery;
- increasing the share of GBS does not mean using it as the only aid instrument; crowding out private and non-governmental stakeholders is a real concern, and a balanced set of aid instruments is a safeguard against that;
- increased donor coordination should not be at the expense of valuable competition in policy ideas (between bilaterals, or between bilaterals and IFIs).

## Evaluation methodology for GBS: issues and challenges

### Box 3: Howard White on lessons from programme aid evaluations

Programme aid = not intended to fund a specific project

Many names, but macro-economically equivalent; systemic effects come from kinds of reporting required

Best 1990s evaluations spanned three questions:

- how has it affected reforms?
- how have reforms affected welfare?
- where has the money gone (in national accounts terms)?

Methods:

- case studies are best, but need to be data-intensive;
- review data and literature first – be prepared;
- but get into the field, use local contacts, hear multiple voices

Discussion moved on to what we have learned from evaluations of programme aid, and what is the relevance to evaluating GBS. Issues debated included:

- what is the right counterfactual? is it with/without? or the same funds using a different modality?
- how firmly can any evaluation questions be answered, given that the counterfactual is always problematic?
- real counterfactuals, as provided by systematic comparisons, versus conjectural ones, as in

writing history – are the second type necessarily inferior?

- can country studies be done with the necessary depth to disentangle the complexity of the causal links? (a lot of data is needed)
- how big is the scope for paired comparisons? (OED has done some at country level, while the joint evaluation of support to basic education implies some at the sector level)

- do we need one logframe for GBS, as in the DFID Evaluability Study, or three covering the three types of input – funds, dialogue and technical assistance?
- how can the effects of GBS be distinguished from those of improved aid harmonisation?

**Box 4: Geske Dijkstra on Dutch debt relief evaluation**

Logframe approach enables:

- a focus on inputs → outputs → outcomes → impacts,
- and thus on efficiency (reductions in debt stock and service flows), effectiveness (improved sustainability, etc.) and relevance (growth and poverty reduction).

The effect on balance of payments and government accounts of Dutch relief cannot be separated from that of other donors. Nor can its effectiveness or relevance be assessed separately.

However, to the extent that Dutch debt relief had stock and flow effects, the general conclusions on effectiveness and relevance are also applicable to Dutch debt relief.

Contributors to the panel discussion suggested:

- GBS evaluation designs need to cover both macro-economic effects and institutional development – we need methodologies

that can do this;

- the long history of methodological debate and learning around the impact of adjustment programmes is still relevant;
- we need prior work on the theory; a joint programme of research and evaluation; and a mix of disciplines;
- recipients must be involved – and national monitoring systems too (but they may be weak where it matters most);
- the development of GBS is at quite an early stage – the required donor coordination, and shared vision between bilateral/IFIs, is still lacking – and this must be taken into account;
- it is questionable whether the immediate priority is full evaluation or refining the logframe and associated theory, and assessing

**Box 5: Tony Killick and Charles Harvey on lessons from SECO evaluation**

Evaluation questions – choosing the right level of ambition: was country X well-chosen as a recipient of GBS? (most basic)

- is GBS having the advantages claimed?
- has it produced good results in country X?
- has *our* GBS produced good results? (best not asked at all)
- has it been more effective than alternative uses - SWAps or projects?

Even the most basic question need lots of information, about political and policy processes, state institutions and capacities, and other donors.

The SECO evaluation suggested:

- GBS can work well, where conditions are right, by shifting emphasis from conditionality to dialogue;
- but this requires heavy-weight economic advisers in-country, better cooperation with IFIs in-country, and more efficient donor coordination and division of labour;
- it is easier to track effects on reducing poverty of public services, than on improving private-sector responses and incomes.

- whether linkages can be tested more rigorously and results measured;
- if evaluation is wanted now, there are two main pathways to explore: 1) concentrate on where paired comparisons are feasible (e.g. across sectors supported in different ways), and 2) address the still obscure attribution and reverse-causality issues by doing more intensive country studies;
- research and theory can help simplify evaluation questions, enabling theory-based evaluation; at the same time, evaluation does not need to wait until all the theoretical issues are clarified;
- a multi-method approach may be useful, as in OED's HIPC Evaluation – not just data-intensive but also interpretation-intensive (any evaluation in this area needs to be theory-rich);
- transaction costs for recipients needs to remain a central question, even if they are difficult to measure – and evaluation designs must be careful not to contribute themselves to this problem;
- in evaluating budget support, we must be sure to include the hidden, political and strategic, objectives – and also the sheer impulse to spend;
- sophisticated methodological innovations such as the application of game theory to Dutch EU negotiations, may not be applicable to less formal relationships and bargaining; but innovations such as focus groups and intensive studies of how organisations change over time are clearly relevant.

**Box 6: Andrew Lawson and David Booth on the GBS logframe and priority issues**

GBS input package includes: funds, policy dialogue and linked TA

Need to focus first on intermediate institutional results in GBS logframe:

- empowerment of government;
- less transaction costs, more allocative efficiency; more predictability;
- increased effectiveness of state;
- greater democratic accountability.

Evaluability Study found “programme intent” less clear than expected, especially among donors.

Inputs are in place and links to immediate results are plausible in Uganda – except that enhanced predictability not being delivered.

For better results, greater clarity is needed on the operational framework:

- GBS = neither traditional conditionality, nor yet pure “partnership”;
- rather a multi-stakeholder club,
- based on interdependency of “loyalty”, “voice” and “exit”.

Implications: further testing and improvement of logframe; work on framework to get more shared vision?

## What are the priority evaluation questions?

Small-group work generated a rich menu of questions about GBS that participants would want to prioritise. In discussion, evaluation specialists commented that these lists distinguished poorly between evaluation questions and research questions, and confused simple and more advanced issues. However, many could be reformulated in more appropriate terms.

The following revised listing reflects some of these comments. But it should be considered raw material for future joint evaluation work, not agreed proposals.

### ***“Four priority evaluation questions”***

- What are the effects of GBS at the level of immediate and intermediate results (i.e. what happens to the money + effects on transaction costs + intended institutional changes/domestic accountability)?
- What are the effects of GBS at the levels of outcomes and impacts, e.g. links from service delivery to effects on private and household sector?
- What are the advantages and disadvantages of GBS and other kinds of support, at a sectoral level in comparable circumstances?
- Under what circumstances will GBS be better practice than equivalent amounts of aid by other means?

### ***“Five priority research topics”***

- Predictability: What is it? How important is it? What are the main sources of it?
- How robust and applicable is the loyalty-voice-exit model?
- What is the potential of national-level monitoring systems?
- Can we collect better data on aid expenditure by modality?
- How much is now known about on the role and effectiveness of ex-ante policy conditionality?

#### **Box 7: Paul Balogun on joint evaluation and the CDF**

A large and important experience, with some useful lessons:

- small management group worked well to get country engagement, large steering group less effective;
- to get learning value, you need to participate;
- use email, video conferencing and websites effectively
- do thematic studies first, using grey literature and the right people;
- formulate questions after agreeing what is already known;
- don't miss out the donor headquarters dimension;
- country studies – generalising from them is hard;
- expect the “so what?” question – to be considered relevant, you have to keep up with current debates, be able to hit a moving target.

## **How might a joint programme of GBS evaluation be organised?**

Various examples of joint evaluation were set out:

- several donors evaluating a multilateral institution jointly;
- UNDP with BMZ on decentralisation;
- joint SPA missions on the Open General License phase of balance of payments support;
- joint country assessments, sometimes including the country;
- joint sector evaluations – e.g. transport in Ghana;
- Rwanda humanitarian intervention and CDF evaluations;

- multi-country basic education support evaluation;
- 8 year evaluation of ACP aid by EU;
- joint synthesis work on research topics – e.g. on peace-building for Utstein Group.

An alternative to all of these is a “coalition” approach in which different agencies undertake different tasks within a common framework. There is then a dilemma about how tight to keep the framework. Partner countries can be included, but first you need a method for choosing them.

The next question is how to manage the evaluation. A steering committee plus a lead agency is a common pattern. Tradeoffs also involved in the decision whether to use a consultant or a consortium, and how much of the design to delegate.

The CDF evaluation may be one of the closest models, because it was a process evaluation and not an impact evaluation, where the problem of the counterfactual becomes more crucial. The CDF experience (Box 7) suggests there may be tradeoffs between gains from joint working and tightness of the framework for synthesising results – e.g. allowing stakeholders in-country to influence the agenda versus a conceptually rigorous agenda.

The question was posed: How would the DAC Working Party take this forward – inventories of existing studies? evaluation guidelines? common terms of reference for country evaluations? commissioned work on thematic issues? research? a learning network? Not all of these suggestions were immediately appropriate, and different avenues were worth considering.

All these questions would need to be considered at forthcoming meetings and bilaterally, taking into account existing efforts such as the Learning Network on Programme-based Approaches, and the SPA. The particular role of the DAC WP in relation to participating agencies and others would need to be considered. DFID saw an early workshop in east Africa to validate work done so far as a useful first step in building a possible evaluation coalition.

## **Some conclusions**

No formal conclusions were adopted, but the facilitation team believe considerable agreement was reached on the following:

- GBS is part of a programme-aid family and has a history – much of what has been learned still applies.
- There are two kinds of effects – macro-economic from the resource transfers, and policy/institutional effects. While the former remain important and need evaluation, the institutional change agenda has been broadened by the current form of GBS.
- We still have limited experience and few studies, but the signs are that when conditions are right (country context + donor behaviour) GBS can work well.

- But there is still much to do, in evaluation terms, and it is not straightforward because of:
  - counterfactual problems;
  - deciding where to focus within the logframe;
  - attribution problems;
  - developing methodologies that are fit for the agreed task.
  
- On the principle that joint financing should always imply joint evaluation, strong efforts are needed to get GBS evaluation onto a solid, collaborative basis.
  
- The role of the DAC WP in this will be critical.

Workshop facilitation team (ODI)  
21 March 2003