Impact evaluation of new aid instruments and country programs

Discussion paper (second draft, December 12, 2007).

Contents:

1 Introduction

Since the mid 1990s, development agencies started to move from project aid towards sector and general budget support. These new aid modalities emerged because of a perceived lack of efficiency and effectiveness of the project approach, due to fragmentation and a lack of coordination, ownership and sustainability. The new aid policy has important implications for the evaluation of the effectiveness of bilateral support. There is quite a lot of experience with doing impact evaluations of specific projects and programs. There is less experience with conducting impact evaluations in the context of sector and general budget support. Rigorous methods for evaluating impact are designed for projects rather than for the evaluation of a sector as a whole. There is a growing consensus on the methods and techniques that can be applied to analyse the effectiveness and impact of specific projects (see the paper of sub-group 1), but there are no general rules for the measurement of the effectiveness and impact of new aid instruments. However, the challenges are not entirely new, since the new aid instruments are to some extent evolved from the policy-based aid which started in the 1980s, so lessons can be drawn from experience in evaluating those programs.

In general, one may discern several approaches for the evaluation of the effectiveness of general budget support and sector support:

1. *Macro studies* try to estimate the effect of aid at the macro level, that is the effect of aid on (for instance) economic growth. Famous are the cross-country growth regressions.
2. Studies that try to assess the effects of aid modalities on the economic and social policies of the recipient countries. The General budget Support study of DFID is an example.
3. *Micro studies* that try to measure the effects and impact of aid on individual households, institutions, and the environment. This paper mentions a few examples of the IEG and IOB. These studies analysis the effectiveness of policies, to which development agencies contribute through program, budget or sector support.

This section of the Guidance addresses specific issues in conducting *quantitative* impact evaluations on sector and budget support and provides guidelines for conducting them effectively. The paragraph on general budget support focuses on macro studies and touches upon a more policy oriented approach. The paragraph on sector support describes a methodology for studies at the micro level. Both sections include several examples. While the value of a mixed-method approach is stressed (see White, 2008), the focus is on the (quantitative) measurement of impact. The section has not the pretension to provide an overall overview of all qualitative and quantitative approaches for the evaluation of program support or to provide a complete toolkit for the evaluation of sector and budget support.

A second warning is in place. It appears to be tempting to compare the effectiveness of different aid instruments: are sector support and general budget support more effective than a more isolated project approach? While this is an important question,

it will be very hard to find the counterfactual. While it is definitely possible to give an *assessment* of the effectiveness of the effectiveness of budgets support or sector support, it seems almost impossible to measure the (quantitative) impact. The problems with (cross-country) growth regressions, included in this section, show that such an approach will be littered with many methodological problems.

Paragraph 2 starts with a brief typology of aid modalities and paragraph 3 discusses the consequences of program aid for the (impact) evaluation of support. The paragraphs 4, 5 and 6 describe several methodologies and instruments for the evaluation of program support. These paragraphs include a few examples as well. Paragraph 7 then gives a few guidelines for managing an impact evaluation. Paragraph 8 mentions possible partners in such an evaluation. The paper ends with conclusions on the methodology. The 5 annexes give some examples and an alternative approach.

2 Aid modalities

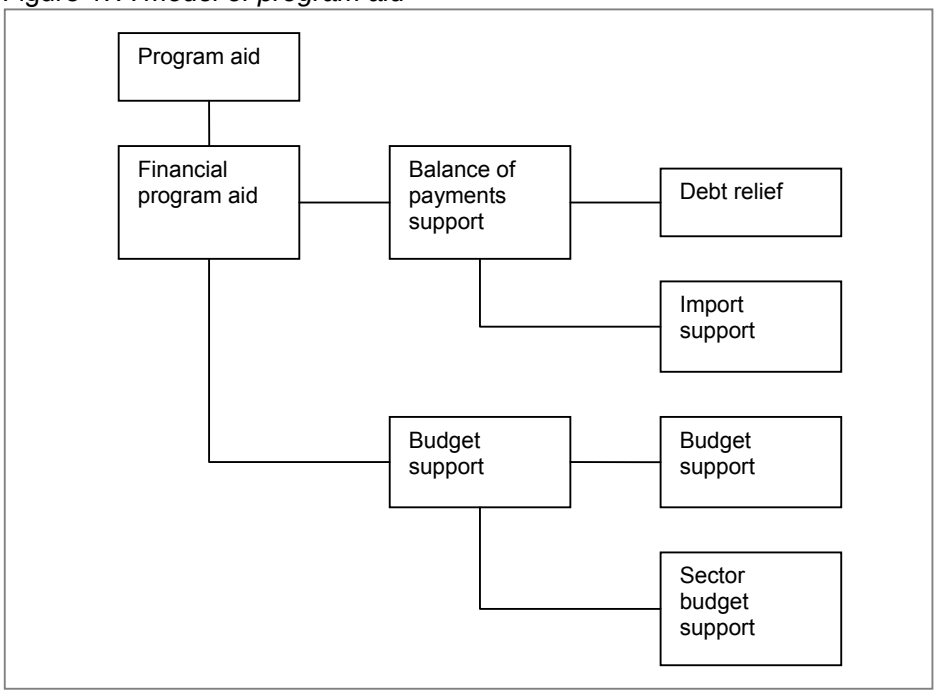
Projects are designed to achieve *specific* objectives within specified resources and implementation schedules. This is no longer obvious in the case of program support. Program aid consists of contributions made available to a recipient country for *general* development purposes and is not linked to specific project activities.

Following White and Dijkstra (2003) four types of program support may be discerned:

1. *Debt relief*: the partial or total cancellation of debts.
2. *Import support*: support to finance the import of goods and services.
3. *General budget support* (GBS): aid to governments that is not earmarked to specific projects or expenditure items. The recipient government can use it to support its expenditure program as a whole. The aid is mixed with the government's own revenues and disbursed through the government's own financial management system.[1]
4. *Sector budget support:* the support to an entire sector based on an agreed policy and expenditure plan for the sector. The aid is earmarked and normally disbursed and accounted for through government systems, preferably with additional sector reporting.[2]
   Moreover, a Sector Wide Approach (SWAp) involves:
   - funding agencies support a shared, sector wide policy and strategy;
   - a medium term expenditure framework and budget to support this policy;
   - government leadership in a sustained partnership;
   - shared processes and approaches for implementing and managing the sector strategy and work program;
   - commitment to move to greater reliance on government financial management and accountability systems.[3]

[1] Definition adapted from the GBS evaluation.
[2] See Foster and Leavy, 2001.
[3] See Walford, 2003.

Figure 1: *A model of program aid*



Adapted from White and Dijkstra (2003) (Note: food program aid, which is of declining importance, is omitted from the original figure).

Ultimately evaluation should be able to answer the question as to the optimal mix of aid instruments. The answer to this question will require a mix of answers on impact as well as process issues (notably on transaction costs) for the range of aid instruments. This end goal is not yet in sight, but to achieve it requires first establishing a firm basis for evaluating new aid instruments, which is the main focus of this paper.

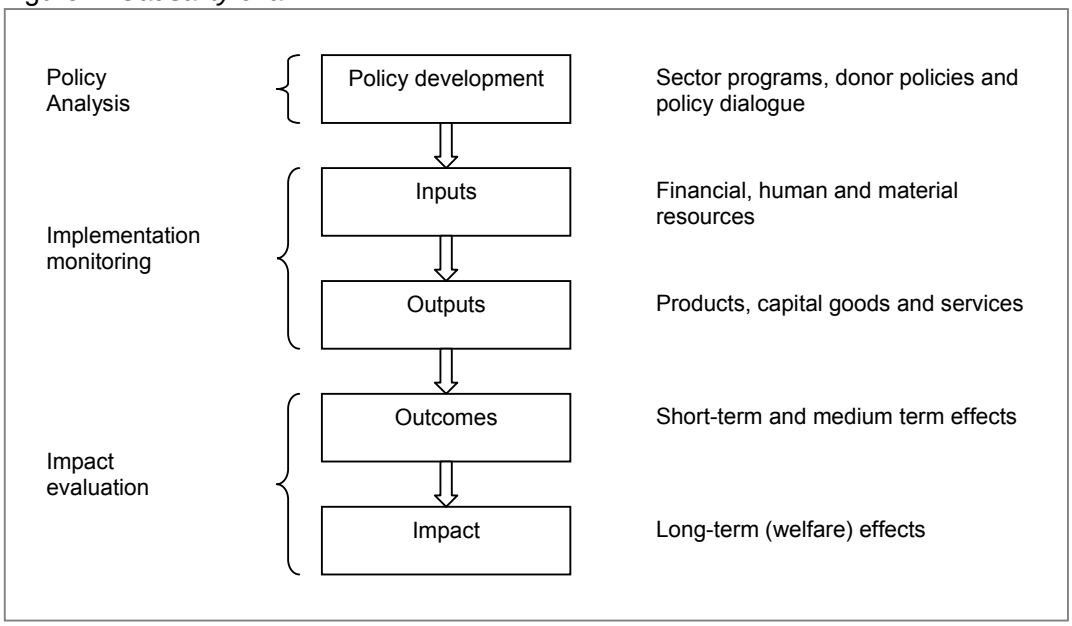
3 The evaluation of budget and sector support

The success of a program depends on the validity of the program theory, the integrity of the sequence and the response of stakeholders on incentives. Any impact evaluation should therefore start with the reconstruction of the program theory. The program theory is a restatement of the log frame which examines how the various links (from inputs, to outputs, outcomes and impact) are meant to work (White and Dijkstra, 2003, p. 20). The core of such a theory is to make explicit the underlying assumptions about how an intervention is supposed to work and then to use this theory to guide the evaluation.[4] The impact evaluation must clarify the logic behind the interventions and the causal (or results) chain from inputs to outcome- or impact-level, and use empirical evidence to verify or invalidate the assumptions which planners and decision makers have based their decision on. The various links in the chain are analyzed using a variety of methods, building up an argument as to

[4] See as well Nicolas Mathieu (2007), Testing Pawson's "Simple Principles for the Evaluation of Complex Programmes" (attached as Annex V).

whether the theory has been realized in practice. In contrast to theory based evaluations, black box evaluations are those which give a finding on impact, but no indication as to why the intervention is or is not giving the expected impacts.

Figure 2: *Causality chain*



Adapted from ADB

The demand for a program theory becomes more compelling when evaluating sector or budget support. These aid modalities redefine the relations between donors and recipient countries. In the case of general budget support or sector support, donors and the recipient country agree on a Poverty Reduction Strategy Paper and a Memorandum of Understanding (MoU). Aid helps the recipient government to implement plans and to achieve its strategic objectives. In addition, there may be an explicit attention to achieve systemic effects; that is to improve the country's financial and budgetary procedures as a result of the systems adopted to manage the program aid funds. A consequence of the move to these aid modalities is an evolution of the causality chain or results chain. GBS and SWAp aim at alignment and increased ownership and within this context the main objective of aid is to help the recipient country to achieve its own objectives.

As a result of this evolution, the evaluation of the effectiveness and impact of aid becomes more complicated. Here Pawson's five principles may be helpful (see Annex 5):

1. Locate key program components. Evaluation should begin with a comprehensive scoping study mapping out the potential conjectures and influences that appear to shape the programme under investigation. One can envisage stage-one mapping as the hypothesis generator. It should alert the evaluator to the array of decisions that constitute a programme, as well as providing some initial deliberation upon of their intended and wayward outcomes.
2. Prioritize among program components. The general rule here is to concentrate on: i) those components of the programme theory which seem likely to have the

most significant bearing on overall outcomes, and ii) those segments of programme theory about which least is known.

3. Evaluate program components by subsets. This principle is about when and where to locate evaluation effort in relation to a programme. The evaluation should take *on sub-sets* of programme theory. Evaluation should occur in ongoing portfolios rather than one-off projects. Suites of evaluations and reviews should track programme theories as and wherever they unfold.
4. Identify bottlenecks in program network. "Theories-of-change" analysis perceives programmes as implementation chains and asks, 'what are the flows and blockages as we put a programme into action?' The basic strategy is to investigate how the implementation details sustain or hinder programme outputs. The main analytic effort is directed at configurations made up of selected segments of the implementation chains across a limited range of programme locations.
5. Provide feedback on conceptual framework. What the theory-driven approach initiates is a process of *'thinking though'* the tortuous pathways along which a successful programme has to travel. What would be described are the main series of decision points through which an initiative has proceeded and the findings would be put to use in alerting stakeholders to the caveats and considerations that should inform those decisions. The most durable and practical recommendations that evaluators can offer come from research that begins with theory and ends with a refined theory.

As for all evaluations, an impact evaluation of sector support and general budget support serves two functions:

a) the *accountability function*. Executive branches of development partners (in donor and recipient countries) need to show results for the continuation of support, the allocation of means for development cooperation and the internal public support for continuing the support. Moreover, impact evaluations are an instrument in the discussions with development agencies on the continuation and the direction of the support.

b) the *learning function*. A well-designed impact evaluation can also answer questions about program design: what works and what doesn't. This should provide policy-relevant information for redesign and the design of future programs.

4 General budget support

GBS seeks to support governments in their response to the needs of their citizens through public investments, by providing services, and by facilitating and regulating the private sector in ways that address poverty, the priorities and modalities (DFID). GBS normally includes:

- a basic agreement between the recipient country and its aid partners about the country's aid strategy, objectives and the principles of cooperation;
- specific agreements about the amounts of budget support and the conditions for its disbursements;
- an agreed procedure for monitoring and review;
- programs to strengthen public finance management.

The impact of general budget support may be measured at several levels. The evaluation may focus on the effect of the support on the policies in the recipient country, on the effect of these changing policies or on the effect of the use of the budgets on households. White (2007) lists four channels through which budget support may have an impact:[5]

- changing policies
- better fiscal management
- the impact of policy changes
- the use of the budget support funds.

All four of these channels are amenable to a degree of impact evaluation using a mixture of methods. An impact evaluation of the first two channels focuses on the impact of donor policy on general economic, social and sector policy in the recipient country. The third and fourth channel are probably related in practice (though not necessarily) and focus on the effects of programs.

4.1 Policy influence

An evaluation of policy influence focuses on the effect of GBS on the government policy in the recipient country and therefore on the upper part of the causality chain. Here, a warning seems to be in place. Many evaluations of donor influence on policy change have relied on simple before versus after comparisons. That is, if policy changes were adopted it was attributed to donor efforts. If they were not, it was most usually put down to 'bad government', or, sometimes, 'political difficulties'. But before versus after is not a good measure of impact – the government may well have changed policies in the absence of donor influences. In general, cooperating agencies may have a tendency to overstate their role in defining policy in the recipient country (see for instance Kanbur, 2000). Some of the literature on conditionality (e.g. Killick et al., 1998) has used statistical analysis, usually single difference, generally concluding that old-fashioned conditionality was not a very successful means of achieving policy change.

A study of policy change therefore cannot usually rely on the sort of tools described above, although there is some statistical literature, both the single difference analyses just mentioned and multivariate analysis of the determinants of policy change which is discussed below. However, in general a case study approach will be appropriate, the evaluation addressing the political economy of policy change and, through a wide range of stakeholder interviews, attempt to locate the donor's actions within that. Moreover, an evaluation should focus on the roles and positions of all parties involved and should include donor behaviour as well as the policy of the

[5] This paragraph and the next draw heavily on White (2007). Large parts are copied from this manual (with the approval of the author).

recipient country. Such an evaluation will be highly *qualitative* (see for instance Grindle, 2007).

A framework for analyzing influence must take into account the following dimensions:

- *actors*: who are the main actors and what are their interests. Why does an actor act the way he does?
- *power relations*: between cooperating agencies and recipient countries and (more important) within the recipient country.
- *Institutional capacity*: what are the institutional constraints (central and local) for the formulation and implementation of government policy;
- *channel*: influence may be formal (conditionality, aid talks etc.), semi-formal (e.g. discussions between senior donor and government officials), and informal (discussions 'in the margins of meetings' or on the cocktail circuit) (see Annex I for examples of influence analysis);
- *direct versus indirect*: a donor may influence the government directly (including through a collective donor effort), or indirectly via: (1) influence on other donors, notably lead donors, or (2) influencing domestic stakeholders.
- *instruments*: there are many instruments for influence. Formal policy dialogue is just one, but others include creating platforms for debate, innovative pilot schemes, technical assistance and financing studies of policy questions and research.
- *directive versus non-directive*: formal conditionality is directive; non-directive policy dialogue facilitates debates without taking sides.

4.2 Better fiscal management

Budget support is accompanied by conditions regarding the management and accounting of the funds. The systems for managing budget support may have systemic effects, whereby government adopts better procedures for all budgetary resources. But there are other reasons why budget systems may change – new government personnel, technical assistance unrelated to the budget support, training programs, or specific IMF interventions. A before versus after analysis cannot establish impact.

Analysis of these systemic effects is the same as that for policy change. First establish a good factual account of before and after. The factual analysis should be available for the monitoring system for public expenditure. Interviews with key stakeholders should be used to identify the sources of any changes in budget management. These interviews should allow respondents the chance to air their own opinions, not leading them to necessarily identify budget support as a causal factor.

A specific aspect of interest is the reduction of transaction costs that arise from the move to program aid rather than project aid. Such an output is certainly amenable to impact analysis, but requires a set of agreed transaction cost measures. Defining such indicators and collecting the required data is a difficult but not insurmountable task. A starting point will be collecting these data on a comparable basis across sectors, across countries and across time. This analysis of the factual will provide a

springboard for analyzing the counterfactual, ideally using a double difference approach.

4.3 The impact of policy changes

Analyzing the impact of policy change was a hot topic in the 1990s as adjustment policies became controversial. The three Reports on Adjustment Lending of the World Bank, a series of reports from IEG as well as independent academic contributions (notably Mosley et al., 1991, and Cornia et al., 1987) are all landmark contributions worth revisiting, although the methods of some would be questioned today (and indeed at the time).

Some policy changes, e.g. health insurance schemes, education vouchers or road pricing, may be introduced on a trial basis in certain areas and so are amenable to 'project impact' methods. Other policy changes, e.g. tariff reduction and tax reform are national so there is no unaffected comparison group. There are several means of conducting an impact evaluation in these circumstances:

1. Cross country studies;
2. Policy modelling using macroeconomic models.

4.3.1 Cross country studies

The first option is to take another country or group of countries as the comparison group. It is of course hard to argue that country X is identical to country Y other than the specific policy of interest. More sophisticated methods were developed by the IMF, and applied by the World Bank, in their analysis of adjustment lending in the 1990s (Goldstein and Montiel, 1986). These methods adjusted for the selection bias arising from the fact that poorly performing countries were more likely to adopt reform by modelling the adoption of reforms. Hence this approach generates a counterfactual policy scenario, and so deals simultaneously with the issues of the determinants of policy change and the impact of policies. Given the renewed concern about selection bias, this approach deserves renewed attention.

In the second half of the 1990s, several authors revived the debate about the effectiveness of aid through cross country-country regressions. The debate started with a study of Burnside and Dollar (2000).[6] According to the authors, aid can be a powerful tool for promoting growth and reducing poverty if – and only if - it is granted to countries with sound development policies and strong institutions. The article provoked an intensive debate. Both authors were criticized, among others by Hansen and Tharp (2001), Easterly (2003), Lensink and White (2000) and Easterly, Levine and Roodman (2004). In 2004 Burnside and Dollar came with new evidence. Others contributed to the debate as well, like Beynon (2003), Rajan and Subramanian (2005) and Masud and Yontcheva (2005). Roodman (2007) gives a summary of the debate, while Rodrik (2005) summarises the problems with growth regressions.

[6] See Collier and Dollar (2002 and 2004) as well.

Following the notation of Roodman (2007), the growth equation of Dollar and Burnside can be written as:

$$\Delta Y = \alpha A + \beta A \times P + \gamma P + \delta X + \varepsilon ,$$

where *Y* is GDP/capita, *A* is aid, *P* is policy, X is a vector of controls, including initial GDP/capita, and ε is the error term. In this specification, the effect of aid on economic growth is a function of aid ánd the interaction of aid and (good) policy. Dollar and Burnside did not find a significant effect of Aid (Aid/GDP), but they did for the interaction Aid x Policy.

When using this kind of approach, it is important to be aware of the problems (see Rodrik, 2005 for a general critique of growth regressions, and White 1992 and Lensink and White 2000 for a critique of their use in the aid-growth literature). Some of them are related to the problems when measuring the impact of projects:

1) the growth regressions are not robust and *sensitive to model specification* and sample selection;
2) *omitted variables and endogeneity*: the aid and policy variables at the right hand side may be endogenous, with the result that the estimates may be biased;
3) *(parameter) heterogeneity*: the regressions presuppose that aid will have the same effect for every country (or that the interaction of aid and policy will have the same effect for every country. Given the large differences between countries, this may be a heroic assumption. Several techniques may be used to deal with country specific effects (like double difference and fixed effects regressions) and they may solve endogeneity problems as well. Nevertheless, these techniques correct for time-invariant differences and will not solve the problem of parameter heterogeneity.
4) *model uncertainty, outliers and influential data points*: a problem with cross-country regressions that the analysis normally includes a limited number of cases, with the effect that outliers and influential data points have a large effect on the estimates. Including or excluding a few outliers may therefore have a large effect on the results. The typical cross country regressions do use panel data (include several years for each country), but this does not necessarily solve the problem.
5) *measurement error:* given the relatively small data sets, measurement error may have a large effect on the results.

Moreover, most studies focus solely on the effect on economic growth thereby neglecting the direct (programme) effects of aid (like for instance the reduction of child mortality) .

A variation on the cross-country regressions gives the IOB study on the effectiveness on international *debt relief* (2003). The study combines a qualitative approach with in depth country case studies with an econometrical approach. The econometrical approach uses panel data for 102 countries. The authors analyse the effects of debt relief on economic growth using the first difference (or changes over time) to get rid of country specific effects (see paragraph 5.2 as well).

4.3.2 Computable General Equilibrium (CGE) Models

Computable General Equilibrium models (CGEs) use a set of equations and actual economic data to estimate how an economy might react to changes, for instance changes in policy. The CGE/ RH models include representative household groups (RHs) and permit analysing simultaneously changes both in the structure of the economy and in the distribution of income. Improved macro-augmented distributional CGERs have been used to analyse the effect of policy changes, for example for the analysis of adjustment policies, on the income distribution and poverty levels in developing countries (Bourguignon, Da Silva and Stern, 2003), and work of the Cornell group specifically on Africa (Sahn et al., 1997).

The building of CGEs is a skill and data-intensive process, but there are CGEs for many countries produced by researchers, so there can be a relatively low marginal cost of utilizing these resources for evaluation purposes (White, 2007). For an overview of the use of CGE models for the assessment of the poverty impact of policies see Khan (2004).

4.3.3 Theories of change

An even more interesting question is how aid and aid modalities affect changes in a society (see Annex V as well). Is there evidence that the changes that the society intended to generate take place? This question addresses the effectiveness of aid in the longer run. What change was expected? What are the markers that tell us we are moving towards that change? What change has happened? What's the evidence that change has happened (who is doing what differently and what if any change in status do we see (inter alia, health, education, welfare status)?[7]

These important questions demand a mixed method approach (see White 2008).

4.4 The impact of budget support resources

Budget support finances government spending, often by pooling resources with other donors. Impact can be analyzed in two stages:

- how the budget support affects the level and composition of government spending;
- the impact of the marginal spending.

It is not possible to trace where an individual donor's "aid dollar" has gone. But this does not mean that nothing can be said about impact. Foster (2004) suggests to analyse general budget support through:

- extrapolation from the budget of the recipient country (if the budget support amounts to x% of the total government budget, then it is fair to assume that x% of the results in sector y have been paid by the budget support);
- the use of notional earmarking;

[7] It is beyond the intentions of this paper to give a broad overview of all theories and factors that have an impact on development effectiveness. It seems nevertheless valuable to give more attention to theories of change in the next version. For a general and broad overview of the impact of aid see Riddell, 2007.

- scoring against those sectors where there is a maximum dialogue (for instance health and education).

The general budget support of the Netherlands to Uganda is an example where the approach of Foster is practical. The budget support is for a large part notionally earmarked to the education sector and this sector is central in the policy dialogue. An education specialist, working on the Embassy of the Netherlands in Kampala, is heavily involved in the sector dialogue.

There is an established literature analyzing the 'fiscal response' to aid (see McGillivray and Morrissey, 2004, for a review). Quite lengthy time series are required for the application of these econometric methods, and there are serious data issues, notably whether aid was on or off budget. These methods may not be applicable. More ad hoc analysis of the sources and uses of funds might nonetheless yield insights.

The second step depends on impact analysis results of government programs. A meta-analysis can be made using existing impact evaluation studies, remembering that an important stage in meta-analysis is quality screening so that only studies of an acceptable degree of rigour are included. Indeed, one of the systemic effects of budget support might be to support more programs amenable to impact evaluation and to ensure quality standards for those evaluations.

This paragraph includes three *examples* that combine, though in a different way, an analysis of the effect of aid on government policies as well as the impact of these (changing) policies.

4.4.1 Country level analysis

White and Dijkstra (2003) analysed the impact of Swedish program aid. Their analysis accepted right from the start that it is impossible to separate the impact of Swedish money from that of other donors' money. Therefore, the analysis focuses on all program aid with nine (country) case studies which trace how program aid has affected macro-economic aggregates (like imports and government spending) and (through these indicators) economic growth. The authors discern two channels for influencing policy: money and policy dialogue. The main evaluation questions are:

1. How has the policy dialogue affected the pattern and pace of reform (and what has been the contribution of program aid to this process)?
2. What has been the impact of the program aid funds (on imports, government expenditure, investment etc)?
3. What has been the impact of reform programs?

Their analytical model treats donor funds and the policy dialogue as inputs, specific economic, social and political indicators as outputs and the main program objectives (like economic growth, democracy, human rights and gender equality) as outcomes and poverty reduction as the overall goal.

The analysis focuses on marginal impact and uses a combination of quantitative and qualitative approaches (interviews, questionnaires and e-mail enquiries). The

analysis of the impact of aid is largely quantitative, while the analysis of the impact of the policy dialogue is mainly qualitative. The authors analyse the impact in three areas: external account, internal account and government finance, where:

External balance:
$AID = M + DS - OKI - \Delta R - X$

Internal balance:
$AID = I + DS - OKI - \Delta R - S$

Disaggregated internal balance:

$$AID = G + I + DS - OKI - \Delta R - T - S$$

Where M= imports, DS= debt service, OKI= other capital inflows, ΔR= change in reserves, X= exports, I= investments, S= (private) savings, G= government expenditure and T= government revenue. This accounting approach is used to identify aid impact on expenditure levels and patterns using a number of ad hoc techniques, such as analyzing behaviour during surges and before versus after breaks in key series, searching the data for other explanations of the patterns observed.

Moreover the authors analyse the impact of aid on stabilization through:
a) the effect on imports;
b) its impact on the markets for domestic currency and foreign exchange;
c) the reduction of inflationary financing of the government deficit.

4.4.2 Joint Evaluation of General Budget Support

An example of a predominantly *qualitative* approach is the Joint Evaluation of General Budget Support (2007). For the GBS evaluation Lawson and Booth (2004) developed a specific evaluation methodology. The final report includes an enhanced evaluation framework (Joint Evaluation of General Budget Support, 2007). The methodology includes standard DAC criteria of relevance, effectiveness, impacts and sustainability and uses a logical framework (or causality map) to analyse the sequence of inputs, immediate effects/activities, outputs, outcomes and impacts. The role of the causality map is to highlight hypothesised causality links (see Annex II). Through this causality chain, GBS is considered to have three types of effects:
a) flow-of-funds effects (macro-economic and budgetary effects);
b) institutional effects;
c) policy effects.

The methodology for determining the impact involves broad *assessments* and not quantitative estimates. The study included case studies for Burkina Faso, Malawi, Mozambique, Nicaragua, Rwanda Uganda and Vietnam.

The method could not solve the attribution problem, though the approach could clarify and disaggregate hypotheses about effects. The evaluation note concludes

that for future evaluations it may be useful to identify components of the evaluation framework that can be evaluated separately.

4.4.3 Total ODA

The *proposal* for a total ODA study combines a qualitative approach with a quantitative impact evaluation (Bigsten, Gunning and Tarp, 2006). In their proposal, the authors discern two channels for the effect of aid on outcomes: a) aid provides financing and b) aid interacts with recipient country policies as well as with policy implementation and institutions. This distinction is comparable with the three effects that have been discerned in the GBS study. The authors suggested to focus the study on the effect on changing policies and the effect on changing policy implementation. The proposed method is that of country case studies with two stages. The first is an analysis of policy choice and implementation and the second stage involves a broad-based application of statistical impact evaluation (see paragraph 5). The second part involves:

1. the identification of a representative sample of government activities and investigation of the availability of data;
2. application of statistical impact evaluation techniques to all the activities in the sample;
3. aggregation of the results.

In practice, an impact evaluation of one sub sector may already be demanding. Therefore another approach *might* be an assessment of the effectiveness of total support based on existing impact studies (at the sector level). With an increasing number of impact evaluations such a meta-evaluation may become more viable.

5 Sector budget support

5.1 Introduction

Many of the same ideas for evaluating budget support apply to evaluating a Sector Wide Approach (SWAP). SWAPs are also expected to affect policy, budget processes, policy outcomes and support financing. Two areas of exception from evaluating budget support may be noted:

- *Use of funds*: general budget support is in principle a free resource—that is, the money is not earmarked against any specific activity. The dialogue between donors and partner governments focuses on overall policy and budget priorities, whereas for sector budget support the focus is on sector-specific concerns. The method of analysis is, however, the same as that used in analyzing how GBS affects the level and composition of government expenditure. In the case of a SWAP this analysis may be extended to the sub-sector level.

- *Factual and counterfactual analysis of outcomes*: Development of a good Management Information System (MIS) for the sector should be part of the SWAP. This system should provide a solid factual, which, combined with the

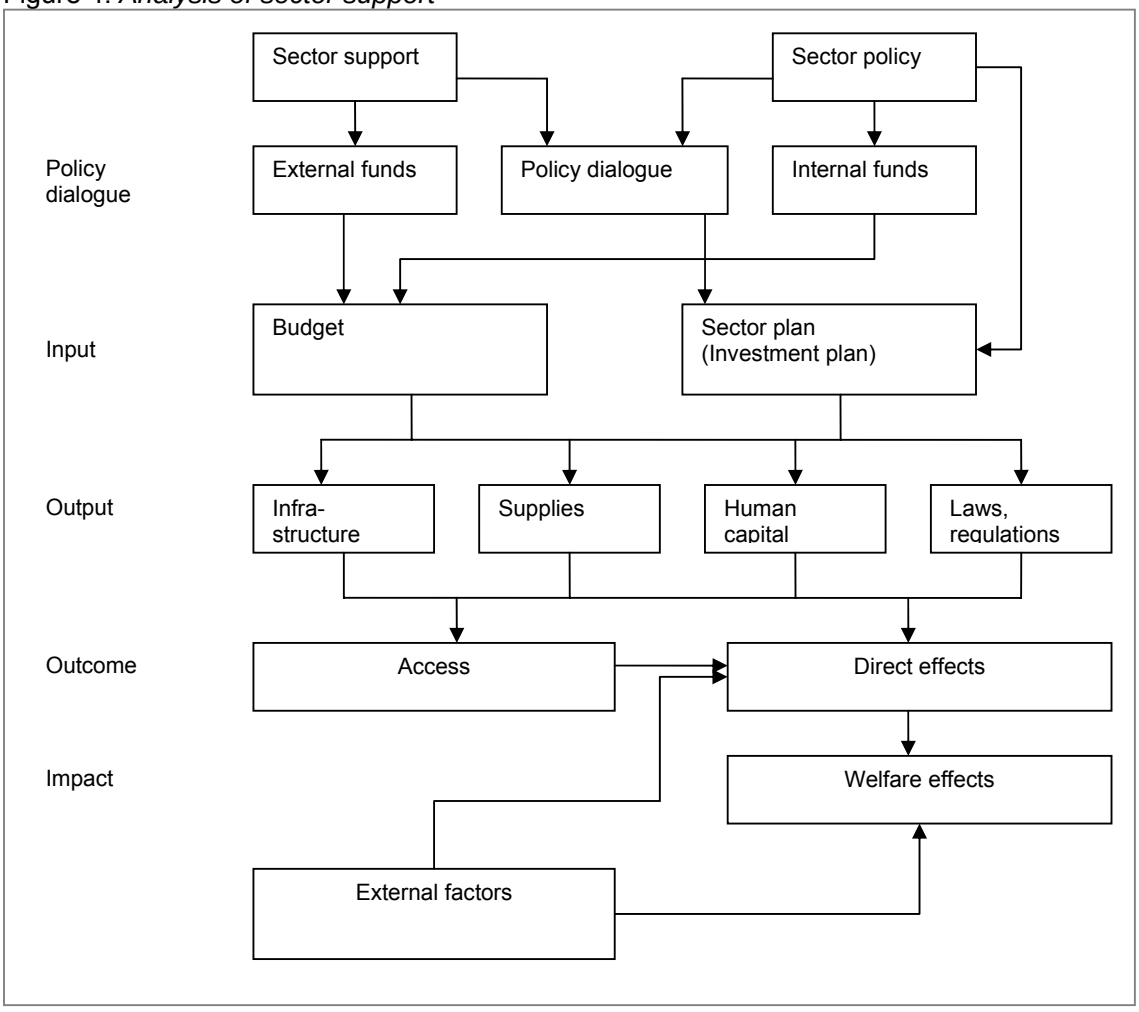
fiscal analysis, makes it possible to *pro rate* what it is that donors have financed. The MIS should also draw on national surveys, such as the Demographic and Health Surveys, to monitor final outcomes. Household surveys are a valuable, and overlooked, source of data on service use. These surveys also validate (or otherwise) the administrative data on matters such as vaccination coverage and give reliable data on facility use. Most importantly, household surveys allow statistical analysis of service use and outcomes, i.e. regression-based counterfactual analysis.

Sector budget support is usually intended to increase funding for the sector of interest. This creates the issue of fungibility. When the government of the recipient country reduces its own spending in the sector in response to the budget support, the effectiveness is low. It should be noted, however, that any foreign involvement may have an effect on the allocation of budgets in the recipient country. Money is fungible, even with earmarking (Bourguignon and Sundberg, 2007). Foster (2004) proposes to ignore the problem. He argues that empirical evidence suggests that earmarking to specific sectors often has little effect on the pattern of spending.

Just like an impact evaluation of general budget support, the evaluation may focus on the sector dialogue and its impact on policies or on the impact of changing policies on outcomes and the final impact. This paragraph focuses on the second approach, this is on the effectiveness of the sector policy*, to which the aid contributes*. This approach analyses how the outcomes of the policy dialogue translate into outputs, outcomes and impact (see figure 4). The impact evaluation focuses on the outcome and impact levels. In this approach, an impact evaluation is an evaluation of the *effects* – positive or negative, intended or not – on individual households, institutions, and the environment *caused by sector-wide interventions*. Such evaluations are not directed at a specific project or intervention, but have the whole sector as object of analysis.

As such, the impact evaluation may be input to or part of a broader evaluation of the progress in the sector and therefore may be part of a mid-term review or an evaluation at the end of the duration of a specific investment plan. Such an impact evaluation is useful when the policy changes and interventions are functioning long enough to have visible effects.

Figure 4: *Analysis of sector support*



5.2 Methodology

A basic requirement of an effective *impact evaluation* at the sector level is the understanding of the methodological problems of an impact evaluation:

- the attribution problem
- selection effects
- the selection on unobservables.

From a methodological point of view, the analysis at the sector level has some advantages as well as disadvantages.

1. A project or program has a clear cut start, but this is not necessarily so at the sector level. This makes it more difficult to measure the before and after situation.
2. In most cases, the intervention is *not discrete and not necessarily targeted at a specific group*. Therefore, it will be difficult or impossible to discern a treatment group and a control group. As a result, it will be difficult to analyse the effect of a policy measure that is implemented at once for the sector as a whole (for instance the abolition of school fees).

3. The *heterogeneity of interventions* makes it more difficult to isolate specific interventions from the rest. Problems of contagion will be all over.
4. On the other hand, a sector analysis may use the *heterogeneity of the objects of analysis*. The evaluation may use, for instances, differences in class size for an analysis of the effects of measures to reduce the class size.
5. In a sector analysis, the use of secondary (administrative) data may be more successful than in a project approach. As a result, there is a greater chance to get panel data. The use of secondary data will define the level of analysis and therefore restrict the choices the researcher has in the selection of the most appropriate unit of analysis.
6. Heterogeneity and possible problems with the quality of data may result in higher error margins. The evaluation may therefore require large samples.
7. When the researcher is able to use secondary (administrative) data, he/she has the advantage of large samples, but may be faced with unlikely figures. Therefore, it is important to analyse the plausibility and reliability of data. There are several ways to do this check:
   a. combining different data sets;
   b. a plausibility analysis (for instance on pupil teacher ratio's, the number of repeaters in comparison with the total number of pupils, etcetera);
   c. an analysis of changes in time.

One of the problems with the analysis of sector support (and budget support) is that normally it is not possible to create a control group. Elbers, Gunning and De Hoop (2007) propose to exploit the heterogeneity within the sector instead. For instance, an analysis of the effect of the building of new classrooms may be analysed by utilising the differences between schools in pupil classroom ratios. The proposed methodology involves a regression of outcome (or impact) variables on the intervention variables:

$$Y_{it} = a + bP_{it} + cX_{it} + \mu_i + \varepsilon_{it}$$

Where Y_{it} is the outcome or impact variable, P_{it} is a vector of policy variables (or interventions) and X_{it} is a vector of control variables. The subscript i denotes the unit of analysis (for instance a school, a pupil, a household or community) and t the time (year) of observation. The authors split the disturbance term in a fixed effect (μ) and a random effect (ε_{it}). The fixed effects are for instance specific characteristics of the individual schools that are not included in the equation. This may be for instance specific characteristics the style of school management or the access to specific resources). The fixed effects may invalidate the results of the regression analysis especially when they are correlated with one or more intervention variables. A well known example is the correlation between class size and the socioeconomic background of pupils. In that case the intervention variable is endogenous. The regression analysis may want measure the effect of differences in class size, but does in fact measure the effect of differences between pupils. Elbers et al propose to deal with the fixed effects by taking the first difference of the original (level) equation:[8]

[8] This equation is analogous to a double difference estimation (see the paper of subgroup I).

$$(Y_{i1} - Y_{i0}) = a + b(P_{i1} - bP_{i0}) + c(X_{i1} - X_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

Here $(Y_{i1} - Y_{i0})$ denotes the change from year t-1 to year t. This may be for instance the change in income, or a change in the number of pupils, a change in the number of teachers, etcetera. Fixed effects are constant in time and as a result they drop from the equation. In the example, the equation measures the effects of changes in class size within schools, rather than comparing schools with different class sizes.

For the variables that do not vary in time the value in year t is the same as the value in year t-1 ($X_{i1} = X_{i0}$)and therefore $X_{i1} - X_{i0} = 0$.

A condition for the use of the proposed technique is that the measurement error is not too large. The method of differencing is more vulnerable to errors in the data than a level regression. In that case, the analysis may lead tot the conclusion that the intervention is not effective, while it actually is.

5.3 Data

For an evaluation of sector-wide support, *secondary data* provide an important source. Moreover, the use of secondary data gives more legitimacy to the gathering of management information. Some of the more common types of secondary data include:

- *administrative data* collected by line ministries and other public agencies (school enrolment, test and examination results, use of health facilities, market prices for agricultural produce);
- *financial data* collected by line ministries and other public agencies;
- national (population) *census data*;
- general *household surveys* such as Living Standards Monitoring Surveys (LSMS);
- *specialized surveys* such as Demographic and Health Surveys (DHS);
- studies conducted by donor agencies, non-government organizations and universities;
- mass media (newspapers, television documentaries, etc). These can be useful, among other things, for understanding the local economic and political context of each project location.

The data gathering process starts with an inventory of available data on the basis of the intervention logic or results chain. On the basis of the program theory and the available data, one may decide on the appropriate level of analysis (households, pupils, schools, communities, etcetera). The next step is an analysis of the quality and reliability of available data, one has to make an inventory of missing data and decide on the need for and design of additional surveys. A pilot must be part of additional surveys. It is necessary to take enough time for the pilot itself and the thorough analysis of the results for the pilot.

5.4 Examples

5.4.1 Sector surveys

A first candidate for analyzing the effects and impact of interventions at the sector level is the use of surveys. An example of such a study is the IEG study on education in Ghana (2004, see Annex III for a detailed summary). There was no clearly defined 'project' for this study, rather World Bank support to the sub-sector through four large operations. The four projects had supported a range of activities, from rehabilitating school buildings to assisting in the formation of community-based school management committees.

To identify the impact of these various activities a regression-based approach was adopted which analyzed the determinants of school attainment (years of schooling) and achievement (learning outcomes, i.e. test scores). For some of these determinants – notably books and buildings – the contribution of the World Bank to better learning outcomes could then be quantified. The analysis of the political economy of education reform in Ghana was a vital piece of the story. The methodology thus adopted a theory-based approach to identify the channels through which a diverse range of interventions were having their impact and had much attention for the context.

As a baseline, the IEG used the *Ghana Living Standards Survey* (GLSS 2). In 1988/89 Ghana Statistical Service (GSS) undertook the second round of the Half of the 170 areas surveyed around the country were chosen at random to have an additional education module, which administered math and English tests to all those aged 9-55 years with at least three years of schooling and surveyed schools in the enumeration areas. Working with both GSS and the Ministry of Education, Youth and Sport (MOEYS), IEG resurveyed these same 85 communities and their schools in 2003, applying the same survey instruments as previously. In the interests of comparability, the same questions were kept, although additional ones were added pertaining to school management, as were two whole new questionnaires – a teacher questionnaire for five teachers at each school and a local language test in addition to the math and English tests. The study thus had a unique data set – not only could children's test scores be linked to both household and school characteristics, but this could be done in a panel of communities over a fifteen year period. The test scores are directly comparable since exactly the same tests were used in 2003 as had been applied fifteen years earlier.

5.4.2 Cross country data and standard surveys

A second example is the IEG study on health in Bangladesh (2005, see Annex IV for a summary). The IEG study utilized existing data sets. The analysis drew on both cross-country data, from a variety of sources, and national data mainly from the Demographic Health Surveys of 1992/93, 1996/97 and 1999/00. Using data of the Demographic and Health Surveys (DHS), IEG contested the widely-held view that fertility decline had reached a plateau in the 1990s. These rates of progress mean that Bangladesh is on track to meet the Millennium Development Goals. Malnutrition

remains high but has begun to decline in the last decade. IEG (2005) examined the factors underlying this success. Multivariate analysis of the determinants of health and nutrition outcomes was carried out. This approach allowed the identification of interventions in a range of sectors which had affected health outcomes. While it was possible to carry out a cost-effectiveness analysis, a full theory-based approach could not be applied because of the absence of process indicators.

5.4.3 The use of secondary (administrative) data

In 2006, IOB has started two impact evaluations of sector support. The first is an evaluation of primary education in Uganda and the second an evaluation of primary education in Zambia. Both studies have been carried out in close cooperation with the ministries of education. Especially in the Uganda study, officers of the Ministry of Education and Sports were involved in the study. Both studies started from the premise that an analysis of the impact of sector support is an analysis of the impact of the policy *to which a specific donor or cooperating agency contributes*. Both studies relied on the use of secondary (administrative data), though they included some field work as well.

The two studies have tried to solve the methodological problems in the following ways:

1. a description of the factual;
2. elaboration of the intervention theory;
3. a statistical, regression based, analysis;
4. triangulation;
5. the exploitation of natural restrictions.

Both studies started with an analysis of developments in the sector based on the literature, documents of the ministries involved and the inclusion of local consultants and experts from the ministry and related agencies in the research team. Next came the development of the *intervention theory* on the basis of the literature (see the IEG study as well). The intervention model relates investments in school management, teachers, classrooms, books and other facilities to specific outcomes as access and learning achievements.[9] Moreover, it includes specific school characteristics as well as pupil, parent and regional characteristics in order to deal with selection effects. The model contributes to a focused theory based data gathering and assures that the statistical analysis is based on the testing of hypotheses rather than on a kind of data mining. The studies use a regression based approach, with the interventions or outputs at the school and other school characteristics as well as pupil, household and regional characteristics as regressors and access and learning achievement as the dependent variables. The main unit of analysis is the school.

The studies are mainly based on the analysis of secondary data. The main sources are:

- annual school census data (for the years 2000-2005);
- national assessment tests;
- examination data (at grade 7);

[9] For further details see: IOB/AIID, 2008.

- Demographic and Health Surveys (DHS), especially the EdData Surveys;
- the Population and Housing Census (2000/2002);
- SACMEQ II data;[10]
- specific surveys for these studies (including a survey on teacher absenteeism).

The heart of the analysis consists of linking EMIS data (school census data in the Education Management Information System) to test and examination results at the school level. The evaluations use DHS surveys and Population Census data for household and regional differences. With school management data for two districts it was possible to analyse the effects of effective management (in Uganda). The linking of data created the possibility to test the reliability of the data. For instance the plausibility of enrolment figures (at the school level) was tested by comparing these figures in time, as well as comparing the enrolment figures at grade 7 with the examination data at grade 7. Other tests checked the consistency of data (by comparing the variables with each other, for instance pupil teacher ratio's, the number of repeaters in comparison with the total number of pupils, etcetera). Extreme figures were deleted from the analysis. First of all, they are probably inappropriate and second, even when they are correct, they are not the main interest of the study. Deleting them reduces the risk of including influential data points at the extremes, that have a relatively large effect on the results.

The use of secondary data has some disadvantages as well. Not all school data appear to be reliable and this may contribute to high error margins. This reduced the possibility to use the techniques of instrumental variables or 'double difference' to get rid of unobserved selection effects. Nevertheless, the robustness of results has been checked through triangulation and the exploitation of natural restrictions. Finally, both studies have been using the technique of *propensity score matching* to analyse the effect of specific (regional) programs.

5.4.4 Conclusion on data

Whilst data availability can remain a problem, the situation has changed markedly in the last twenty years. Most African countries now have at least two, and some many more, income and expenditure surveys, and a good number are in their third and fourth round Demographic and Health surveys. The existence of these data provides two opportunities. The first is to conduct ex post impact evaluations exploiting existing data. The second is to incorporate analysis of these data into sector and national M&E systems. Doing so will generate some user demands as to the nature of the data (e.g. linked facility surveys) which will increase their utility for impact evaluation purposes. Whilst in the short term external agencies may need to get some studies done, in the medium to long-term the focus should be on building domestic capacity to conduct such policy-relevant analysis within government systems.

[10] SACMEQ is the Southern and Eastern Africa Consortium for Monitoring Educational Quality. The consortium (of fifteen ministries of Education) tests the progress of pupils on reading and mathematics.

6 Additional instruments

6.1 Poverty mapping

Specific household surveys like the demographic and Health Surveys (DHS) or the Living Conditions Monitoring Surveys (LCMS) contain a wealth of information for assessing the impact of specific programs (see for example the IEG study on health in Bangladesh). A disadvantage of the surveys may be the relatively small sample size. This may be especially a hindrance for the analysis of the impact of projects, though the sample size may be to small for the analysis of country wide programs as well (depending on the unit of analysis).

Census data (of the Population and Housing census) provide information at the household level, but in general this is not a very rich database and does not contain information on (for instance) the income of households. The technique of poverty mapping consists of matching census data with data from household surveys. The technique permits to estimate the mean and distribution of household characteristics that are not directly observed in household surveys. The technique uses the smaller and richer survey to estimate the joint distribution of a household variable that defines the distribution of poverty or inequality. By restricting the set of explanatory variables to those that can also be linked to households in a larger sample or census, this estimated distribution can be used to generate the distribution of this household variable for a subpopulation in the larger sample conditional on the observed characteristics (see Elbers, Lanjouw and Lanjouw, 2002). The advantage of poverty mapping is an efficient use of data from different sources. It combines information from extensive surveys (though with relatively small sample sizes) with census data. The technique derives estimators of welfare from the relatively small household surveys and apply the estimates on census data.

The data of a poverty map may be used to include specific household characteristics in the impact evaluation. For instance, in the (IOB) Zambia study, the results of a poverty mapping procedure have been used to analyse the relation between welfare level (poverty level) and the allocation of resources for schools.

Depending on its specific purpose, census data may be used directly as well. For example, census data include information on the level of education and the job status of the adults in a household. In the (IOB) Uganda and Zambia studies, this information has been used to construct a social-economic variable at the local level (Parish level for Uganda and Ward level for Zambia). With the method of principal components analysis, several variables of the household survey were used to construct one new variable.

6.2 Tracking surveys

An impact evaluation may be combined with a Public Expenditure Tracking Survey (PETS). Such a survey may traces leakages in the allocation of funds. A first PETS for Uganda in the mid 1990s (with data for 1991-1995) showed that only 13 percent of non-wage recurrent expenditures for primary education actually reached the

primary schools (Reinikka and Svensson, 2001). Moreover, larger schools and schools with wealthier parents received a larger share of the intended funds (per student) than smaller schools and schools with poor parents. The variation in grants received across schools was determined more by the political factors than by efficiency and equity considerations.

A PETS tracks the flows of resources to determine how much of the originally allocated resources reach each level (Dehn, Reinikka and Svensson, 2003). It locates the leakage of funds and provides information on the actual allocation. A related instrument is a Quantitative service Delivery Survey (QSDS). Such a survey analyses the actual service delivery. Its primary aims are to examine the efficiency of public spending and to analyse the incentives for frontline service providers to deliver.

Both tools recognise that a service provider may have a strong incentive to misreport. PETS and QSDS deal with this problem through a multiangular data collection strategy, combining data from different sources. This strategy serves as a means of cross-validating the information from each source. A PETs includes:

- the identification of key service delivery issues and problems;
- determination of the structure of the resource flows;
- a good understanding of the institutional setting;
- a check on data availability;
- a survey strategy and sampling procedure;
- questionnaire design;
- an assessment of the local capacity to carry out the survey.

7 Managing and implementing impact evaluations

1. The (proposal for an) evaluation should begin with a comprehensive scoping study mapping out the potential conjectures and influences that appear to shape the program under investigation. Establishing the program theory can (should) be a participatory process. This process may be a useful part of the evaluation findings as the program theory can make explicit the logic of the interventions.

2. The proposal must be discussed with the ministry of the recipient country involved and with other development partners. The writing of the (draft) ToR and discussion of the draft with the recipient country and other development agencies and finalising the draft on the basis of inputs from the other partners may take up to *three months*. This phase already includes a visit to the ministry and other agencies active in the sector, in order to be able to assess the availability and quality of data.

3. During the discussions, one may decide to do a *multi-donor* evaluation.[11] It is important nevertheless that the evaluation remains manageable and objective and will not be used for political goals of development agencies.

[11] See OECD/DAC, Effective practices in conducting a multi-donor evaluation, Paris, OECD, 2000.

4. The ToR should require that a clear understanding of the intervention be a prerequisite for the evaluation design. Sector and area expertise may not be essential but is certainly an advantage. The ToR for an impact evaluation should also stress the need for a credible counterfactual analysis. Proposals or concept notes should make clear how this issue will be addressed, being explicit about the evaluation approach.

5. The evaluation team must consist of members with the technical competence to assess and solve the specific methodological problems, as well as members with specific knowledge of the sector. When the evaluation will be tendered, this process may take up to another three months.

6. The ToR must include a proposal for the *evaluation period*. This period must be long enough to be able to assess the effects of policy changes or the effects of interventions. For the statistical analysis, it may be possible to select several years within the total evaluation period. It is not always necessary to gather data for every year.

7. For sector and country studies, *establishing a solid factual* is a good starting point.

8. Test, test and re-test the instruments. A well designed pilot is an absolute precondition for a successful survey. One must be sure that the pilot is representative for the problems that may be expected. Run planned tabulations and analyses with dummy data or the data from the pilot. Once data are collected one to two months are required for data entry and cleaning.

9. Good quality data are essential to good impact evaluation. The evaluation design must be clear on the sources of data and realistic about how long it will take to collect and analyze data. The process of analysing the quality, the processing and linking of *secondary data* may take up to three months. Moreover, it is important to include time (three months) for getting the data, even though they should be already available somewhere.

10. The impact evaluation requires that data are collected at a very low level of aggregation. This may be a household, a school, a hospital, etcetera. The management of the evaluation must agree with the ministry of the recipient country and other agencies involved on the conditions for the use of the data (confidential character, at what level of aggregation will be reported). The ministry and other agencies will remain owner of the data.

11. There must be a thorough analysis of missing data and ways to get additional information. An *additional survey* may take up to three to four months. When the results of this survey will be the main source of information, one should reckon with up to six months.

12. The *data analysis* may take three to four months. The researchers will encounter unexpected problems that will have to be solved during that analysis. This may even require additional data.

13. It is recommended to include at least one month of analysis with the team in and with experts from the recipient country. This will enhance ownership, contribute to a better understanding of problems encountered and speed up the process of getting additional data, when needed.

14. The inclusion of several stakeholder workshops may enhance the whole process. The evaluation may include three workshops with local stakeholders: at the beginning, halfway and at the end.

15. The process of discussing the draft report must respect the standard procedures within the ministry of the recipient country and the procedures that the recipient country and development partners have agreed on. This may take up to two months and may include a stakeholder workshop.

16. The time required for an impact evaluation depends on whether primary data collection is involved. If so, 18 months is a reasonable estimate from inception to final report.

8 Evaluation partners

A sector wide evaluation requires the involvement of the recipient country, because the country:
a) is the owner of the sector policy;
b) brings in a broad experience and specific knowledge of the sector;
c) is familiar with the specific problems in the sector;
d) has the network for an effective evaluation;
e) knows what kind of information and data are already available.

The evaluation department that has taken the initiative for the impact evaluation must first of all seek the participation of appropriate representatives from the recipient country's government ministries and agencies. Moreover, participation by recipient country nationals often brings to the evaluation effort special expertise regarding country context and program implementation issues. Another advantage is that participation in the evaluation process also brings about ownership of the evaluation's findings and recommendations, thus making follow-up on actions by the host country institutions more likely.

Capacity building at the level of the ministry involved should be an explicit purpose of the evaluation. Therefore, the evaluation team must include one or more representatives from the ministry of the recipient country and a local consultant, preferably from a university.

In cases where sector-wide investment programs are financed by multi-donor co-financing schemes, the participating donors would make natural partners for a joint evaluation of that sector program.[12] Other factors in selecting other donors as

[12] See OECD/DAC, Effective practices in conducting a multi-donor evaluation, Paris, OECD, 2000. This paragraph is copied from these guidelines.

partners in joint evaluation work may be relevant as well. Selecting those with similar development philosophies, organizational cultures, evaluation procedures and techniques, regional affiliations and proximity, etc. may make working together easier. Another issue may be limiting the total number of donors participating to a "manageable" number. In cases where a larger group of donors is involved, a key group of development partners (including the recipient country) may assume management responsibilities where the role of others is more limited.

Once appropriate donors are identified that have a likely stake in an evaluation topic, the next step is to contact them and see if they are interested in participating. In some cases, there may already be an appropriate donor consortium or group where the issue of a joint evaluation can be raised and expressions of interest can be easily solicited. The DAC Working Party on Aid Evaluation, the UN Evaluation Group (UNEG) and the Evaluation Cooperation Group (ECG) have a long tradition of co-operation, shared vision on evaluation principles, and personal relationships built over the years and have fostered numerous joint evaluations.

9 Conclusions

This paper includes several approaches for the evaluation of impact of ‘new’ aid instruments and country programs. The paper started with a typology of program aid and situated the program aid within a causality chain or (results chain). A result of the new aid instruments is an evolution of the causality chain. This raises new questions for the analyses of the effectiveness of aid. What is the object of analysis? This may be the impact on policy and policy change, on service delivery and on outcomes and impact at the household level.

The analysis of the impact of general budget support and sector support on the government policy in the recipient country normally boils down to a qualitative analysis. This may be a political analysis, using methods and techniques from the political sciences or public administration. Such analyses help to understand political processes and determinants of policy change within countries and (power) relations between donors and recipient countries, but may not increase the understanding of the effectiveness at the other end of the results chain, that is the effects on households. In the end, when it comes to the question of the effectiveness of aid, this approach may prove to be unsatisfactory.

Several studies have tried to estimate the effectiveness of aid through cross-country studies using econometrical techniques (cross-country-regressions). The studies of Burnside and Dollar, and the reactions on their studies, are famous. These techniques may be applied to study the effectiveness of aid, though a warning seems to be in place. The methodological challenges are enormous and require a thorough understanding of the intervention mechanisms and methodological (econometric) problems. Many cross-country regressions have ignored the causality chain and handled the relation between aid and development as a black box (Bourguignon and Sundberg, 2007). Instead of focusing on the cross-country regressions, the authors propose to come up with more impact evaluations that analyse the relation between government policies and country outcomes.

This paper includes a few examples of studies that try to evaluate the effect of general budget and sector support. The comprehensive Joint Evaluation of General Budget Support developed an impressive framework for analysis, but did not analyse the impact of the support with quantitative techniques. Bigsten, Gunning and Tarp (2006) propose a two stage approach that combines qualitative studies with statistical impact evaluation. The third example, the study of White and Dijkstra (2003) combines a qualitative and quantitative approach as well. The main difference with the proposal of Bigsten, Gunning and Tarp is that the quantitative analysis of White and Dijkstra is at the macro level, while the first three authors propose to select several government activities for a (micro) impact evaluation. The examples of sector studies show that this approach may prove to be demanding. An impact evaluation of a sector is already a challenge in itself. A strategy may be to make better using of existing evaluations through a (quantitative) meta-analysis or a (qualitative) meta-evaluation.

While agencies could and should make better use of existing evaluations, this does not preclude the need for more impact evaluations. With the improvement of administrative data, the use of secondary data may increasingly become an option, while at the same time this will lead to a better utilisation of existing data. However, additional surveys will be needed as well.

When conducting an impact evaluation, it is important to find a balance between statistical rigour and a theory based approach, that rests on a thorough understanding of the sector. Researchers may be overwhelmed by the ‘endogeneity curse’ (Thorbecke, 2006, p. 27). The emphasis on combating (possible) econometric biases runs the risk of excluding larger structural and conceptual issues. A balance between a qualitative theory based approach, a contextualised analysis and a rigorous quantitative analysis seems to be the most promising strategy.

Literature

- Bamberger, Michael and Howard White (2007), Using strong evaluation designs in developing countries: experience and challenges, in: *Journal of Multidisciplinary Evaluation*, volume 4, Number 8, pp. 58-73.
- Beynon, Jonathan (2003), Poverty Efficient Aid Allocations – Collier/Dollar Revisited, ESAU Working Paper 2, London, ODI.
- Bigsten, Arne, Jan Willem Gunning and Finn Tarp (2006), *The effectiveness of total ODA: an evaluation proposal*, Göteborg / Amsterdam / Copenhagen.
- Bourguignon, François and Luiz A. Pereira da Silva, (eds.), 2003, *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*, Oxford University Press.
- Bourguignon, François and Mark Sundberg (2007), Aid effectiveness, opening the black box, in: American Economic Review, 2007, vol. 97 (2), pp. 316-321.
- Burnside, Craig and David Dollar (2000), Aid, policies and growth, in: American *Economic Review* 90 (4), pp. 847-868.
- Chauvet, Lisa and Patrick Guillaumont (2002), *Aid and Growth Revisited: Policy, Economic Vulnerability and Political Instability,* Paper presented at the Annual Bank Conference on Development Economics: Towards Pro-poor Policies, Oslo.
- Collier, Paul and David Dollar (2002), Aid Allocation and Poverty Reduction, in: *European Economic Review* 45(1), pp. 1–26.
- Collier, Paul and David Dollar (2004), Development Effectiveness: What Have We Learnt?, in: The *Economic Journal* 114(496), pp. 244–71.
- Cornia, Andrea, Richard Jolly and Frances Stewart (1987) *Adjustment with a Human Face* Oxford: Oxford University Press
- Dehn, Jan, Ritva Reinikka and Jacob Svensson (2003), Survey tools for assessing performance in service delivery, in: Francis Bourguignon and Luiz A. Pereira da Silva, (eds.), *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*, Oxford University Press.
- Dijkstra, Geske and Niels Hermes (2003), *The debt crisis, international responses, and results of debt relief*, The Hague, Netherlands Ministry of Foreign Affairs.
- Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw (2002), Micro-level estimation of welfare, in: *Econometrica* 71:1, pp. 355-364.
- Elbers, Chris, Jan Willem Gunning and Kobus de Hoop (2007), *Assessing budget support with statistical impact evaluation*, Amsterdam, Tinbergen Institute, Discussion Paper, TI 2007-075/2.
- Easterly, William (2003), Can foreign aid buy growth? In: *Journal of Economic Perspectives,* Volume 17, Number 3,pp. 23-48.
- Easterly, William, Ruth Levine and David Roodman (2004), New Data, New Doubts: A Comment on Burnside and Dollar's "Aid, Policies, and Growth" (2000), in: *The American Economic Review*, Vol. 94, no. 3, pp. 774-780.
- Foster, Mick and Jennifer Leavy (2001), *The choice of financial aid instruments*, London, ODI.
- Foster, Mick (2004), *Accounting for donor contributions to education for all*, Essex.

- Hansen, Henrik and Finn Tarp (2001), Aid and growth regressions in : *Journal of Development Economics* 64, pp. 547-570.
- Goldstein, M. and Peter Montiel (198), Evaluating Fund stabilization programs with multicountry data: some methodological pitfalls. *IMF Staff Papers 33* June
- Grindle, Merilee S. (2007), Good enough governance revisited, in: Development Policy Review 25 (5), pp. 553-574.
- IOB (2003), *Results of international debt relief 1990-1999* , The Hague, Netherlands Ministry of foreign Affairs.
- IOB / Ministry of Education and Sports (2008), *Impact evaluation of primary education in Uganda* (draft), The Hague/Kampala, IOB/MoES.
- IOB / AIID (2008), *Impact evaluation of primary education in Zambia* (draft), The Hague/Amsterdam, IOB.
- Joint Evaluation of General Budget Support 1994-20054 (2007), *Evaluation of General Budget Support – note on approach and methods*, London, DFID.
- Kanbur, R. (2000), Aid, Conditionality and Debt in Africa, in: In Finn Tarp (ed.) *Foreign Aid and Development: Lessons Learnt and Directions for the Future.* London and New York: Routledge.
- Khan, Haider A. (2004), Using Macroeconomic Computable General Equilibrium Models for Assessing Poverty Impact of Structural Adjustment Policies, Manila, ADB, ADB Institute Discussion Paper No.12.
- Killick, Tony, with Ramini Gunatilaka and Ana Marr *Aid and the Political Economy of Policy Change*, London and New York: Routledge, 1998
- Lensink Robert and Howard White (2000), Assessing Aid: A Manifesto for Aid in the 21 st Century, *Oxford Development Studies*, vol.28, no.1
- Masud, Nadia and Boriana Yontcheva (2005), *Does Foreign Aid Reduce Poverty? Empirical Evidence from Nongovernmental and Bilateral Aid*, IMF.
- McGillivray, M., and O. Morrissey (2004), 'Fiscal Effects of Aid' in T. Addison and A. Roe (eds), *Fiscal Policy for Development*, Basingstoke: Palgrave Macmillan/WIDER, pp. 72-96. Also as *CREDIT Research Paper 01/13* and *WIDER DP 2001/61*
- Mosley, Paul, Jane Harrigan and John Toye (1991) *Aid and Power*, London: Routledge
- Nguyen, Bin T. and Erik Bloom (2006), *Impact evaluation, methodological and operational issues*, Manila, Asian Development Bank.
- Rajan, Raghuram G. and Arvind Subramanian (2005), What undermines Aid's impact on growth?, IMF Working Paper 05/126, Washington DC, International Monetary Fund.
- Rajan, Raghuram G. and Arvind Subramanian (2005), Aid and Growth: What Does the Cross-Country Evidence Really Show?, IMF Working Paper 05/127, Washington DC, International Monetary Fund.
- Reinikka, Ritva and Svensson, Jakob (2001), *Explaining Leakage of Public Funds*, Washington DC, World Bank Policy Research Working Paper No. 2709.
- Reinikka, Ritva and Jacob Svensson (2004), Local capture: evidence from a central government transfer program in Uganda, in: *The Quarterly Journal of Economics*, vol. 119 (2), pp. 678-704.
- Riddell, Roger C. (2007) *Does foreign aid really work*, Oxford, Oxford University Press, 2007.

- Rodrik, Dani (2005), *Why We Learn Nothing from Regressing Economic Growth on Policies*, Cambridge, Harvard University.
- Roodman, David (2007), *The Anarchy of Numbers: Aid, Development, and Cross-country Empirics*, Washington, Center for Global Development, Working Paper 32.
- Sahn, David, Paul A. Dorosh, and Stephen D. Younger (1997) *Structural Adjustment Reconsidered: Economic Policy and Poverty in Africa*, Cambridge University Press
- Thorbecke, Erik (2006), The evolution of the development doctrine, 1950-2005, Helsinki, UNU-Wider, Research-Paper no. 2006/155.
- Walford, Veronica (2003), *Defining and evaluating SWAps*, London, Institute for Health Sector Development.
- White, Howard (1992) The macroeconomic impact of development aid: a critical survey. *Journal of Development Studies*;
- White, Howard and Geske Dijkstra (2003), *Programme aid and development, beyond conditionality*, London, Routledge.
- White, Howard (2007), DFID Economists' manual, London, DFID.
- White, Howard (2008) Of probits and participation: the use of mixed methods in quantitative impact evaluation, in: *IDS Bulletin* 39.1.

ANNEX I Examples of influence analysis

Channels for influence: the Swedish experience

	Formal channels	Semi-formal channels	Informal channels
Direct influence on government	Formal channels, such as annual aid negotiations and CG meetings, seem to have a relatively limited role insofar as direct influence on government is concerned for individual bilaterals. CGs are, however, important as a focal point for policy dialogue, providing an opportunity for co-ordinated bilateral action. Sometimes disagreements may become formalised, the constitutional crisis in Nicaragua being the most prominent example in which Sweden has played the lead role. In other cases, such as Zambia, Sweden has acted as a part of the donor community. An exception to this general statement is where PA modalities have been a subject of these discussions and are playing a leading role in market development: Vietnam is perhaps the only case. Sweden has not in general adopted own conditionality independent from the IFIs; the exception being the experiment with Matching Funds in Tanzania.	"Presence" determines effectiveness of direct semi-formal links: very evident in Vietnam, and has been an important factor in other countries studied except Bangladesh and, possibly, Cape Verde. Studies and seminars have been used to good effect in several countries (notably Vietnam and also, though not part of this evaluation, Guinea-Bissau), and impact is assisted by presence - though this is not sufficient condition in cases where there is much competing material (as seems to be so for Tanzania). Policy seminars for civil society, as regularly held in Zambia, are a special case here. Semi-formal (and informal) links between ambassador/embassy officials and head of state/ministers are uneven. In cases where presence is high then ambassador has had strong and important links.	Swedish TA supports policy advice through the expertise it provides, but there was not evidence of using TA as an additional channel to exert pressure (indeed there was some opposition to the idea). The margins of CG meetings provide an important opportunity for intensified informal contacts with government officials.

	Formal channels	Semi-formal channels	Informal channels
Indirect influence via IFIs	The main formal channel, the Board, is commonly agreed to be ineffective. Opinions vary on impact *via* SPA, though difficult to find evidence at country level. The content of policy-conditionality has not strongly reflected SPA concerns on poverty or gender. JEMs were important for the management of programme aid, but cannot be held up as an instance of bilateral influence. No evidence was found of influence at the country-level *via* Trust Funds.	Less formal channels, as enquiries by Delegation staff and direct high-level communications do not appear to be much utilised. Contacts with in-country and visiting IFI missions, though often good (e.g. Tanzania and Vietnam - though in other places the IFI Resident Mission is seen as weak) is not seen as a channel for influence.	There are informal contacts from agency to IFI staff, but no instances were found in which these contacts can be said to have had influence. This is a general finding (i.e. applies to other bilaterals) and reflects the culture of the IFIs.
Indirect influence via donor agencies	There are few formal channels through which one bilateral can influence others; aid co-ordination bodies (DAC, SPA, CGs) present some opportunity, but it is not marked. The activities of like-minded groups may sometimes become more formalised	Sweden is often active in like-minded groups, e.g. in Mozambique and the non-group of donors in Zambia. Studies can help inform other donors in cases where there is a dearth of material; Guinea-Bissau is such a case, as was Vietnam in the late 1980s.	Informal links between bilateral agencies are clearly of great importance. In cases where the donor community is close-knit - which is most countries but in some, e.g. Zambia, more than others - then there is a common position which emerges rather than being set by a specific donor. Hence there are few instances of “leadership by example” (but the Nicaragua constitutional crisis is one), but to emphasise this lack would be to understate influence through these channels. Similar comments may be made concerning influence through informal channels associated with formal aid co-ordination, such as the margin of CG and SPA meetings.

Annex II: Evaluation Framework Joint Evaluation of GBS

– Logical sequence of effects

Level 1 (the design)
1. Adequate quantity and quality of inputs are provided by new GBS:
1.1 Funds
1.2 Policy dialogue
1.3 Conditionality
1.4 TA/capacity building linked to
• Public finance management (PFM)
• Pro-poor sectoral policies and good governance
1.5 Alignment and harmonisation
• International Partners' (IPs') alignment to government goals and system
• IPs' harmonisation

Level 2 (the immediate effects/activities)
2.1 More external resources for the government budget (additionality)
2.2 Proportion of external funds subject to national budget process increased (increased fungibility)
2.3 Increase in predictability of external funding of national budget
2.4 Policy dialogue and conditionalities focused on pro-poor policy framework and improved PFM
2.5 TA/capacity building established to:
• improve PFM processes including budgeting, accounting, financial control, audit
• improve the linkage between PFM and pro-poor sectoral policies and good governance
2.6 Actions to ensure IPs' alignment are in place
Actions and agreements to improve IPs' harmonisation are in place

Level 3 (the outputs)
3.1 Increased resources for service delivery:
• External resources are treated as additional
• Cost of funding budget deficit reduced
3.2 Partner government is encouraged and empowered to strengthen PFM and government systems:
• To use the budget to bring public sector programmes into line with government goals, systems and cycles (Poverty Reduction Strategy Paper/Medium Term Expenditure Framework)
• To set up performance monitoring systems to measure the effectiveness of public expenditure at the level of the final beneficiaries
• To promote alignment and harmonisation by IPs
3.3 Partner government is encouraged and empowered to strengthen pro-poor policies:
• To establish and execute an adequate sequence of reforms to ensure macroeconomic stability and private sector development
• To establish and execute pro-poor policies and targeting in health, education, agricultural and rural development
• To enhance social inclusion policies, through decentralisation and participation of the civil society, reform of the administration of justice and respect for human rights
3.4 Improved aggregate fiscal discipline:
• More predictable funding flows
• Incidence of liquidity shortfalls reduced, hence less use of Central Bank overdrafts and less accumulation of arrears

3.5 Operational efficiency of public expenditure is enhanced:
• By reductions in certain types of transaction costs to partner government (e.g., non-standard procurement systems, brain-drain effects of parallel project management structures)
• Better planning, execution and oversight reduces wasteful spending, controls corruption better, spreads positive lessons across the public sector
3.6 Allocative efficiency of public expenditure is enhanced:
• By a more effective budget process: multi-year, results oriented, transparent, participatory; with effective execution and audit; with an adequate tracking system
• By increased capture of project funds in budget
• By stakeholders taking the domestic budget more seriously (because that's where the money is)
3.7 Intra-government incentives and capacities are strengthened:
• Official reporting lines are more respected (vertical through government to cabinet, not horizontal to IPs)
• Public-service performance incentives are strengthened, so that policies are made and implemented, audit and procurement systems work, and corruption is reduced
3.8 Democratic accountability is enhanced:
• Greater role of parliament in monitoring budget results
• Accountability through domestic institutions for IP-financed spending is enhanced
• Conditions for all-round democratisation are thereby improved, including the trust of people in their government and hence their level of expectations

Level 4 (the outcomes)
4.1 Macroeconomic environment is favourable to private investment and growth:
• Inflation controlled
• Realistic exchange rate attained
• Fiscal deficit and level of domestic borrowing sustainable and not crowding out private investment
4.2 Regulation of private initiative works to ensure business confidence, equity, efficiency and sustainability:
• Policies on corruption, property rights resolutely pursued
• Market-friendly institutions developed
4.3 More resources flowing to service delivery agencies
4.4 Appropriate sector policies include public actions to address major market failures, including those arising from gender inequalities
4.5 More effective and accountable government improves administration of justice and respect for human rights, as well as general confidence of people in government
4.6 More conducive growth enhancing environment
4.7 Public services effectively delivered and pro-poor:
• Service delivery targets met for key pro-poor services
• Evidence of increased use of services by poor (including poor women)

Level 5 (the impact)
5.1 Income poverty reduction
5.2 Non-income poverty reduction
5.3 Empowerment and social inclusion of poor people

Copied from: Joint Evaluation General Budget Support (2007), pp. 19-20.

ANNEX III: Basic Education in Ghana

Introduction

In 1986 the Government of Ghana embarked on an ambitious program of educational reform, shortening the length of pre-University education from 17 to 12 years, reducing subsidies at the secondary and tertiary levels, increasing the school day and taking steps to eliminate unqualified teachers from schools. These reforms were supported by four World Bank credits – the Education Sector Adjustment Credits I and II, Primary School Development Project and the Basic Education Sector Improvement Project. An impact study by the World Bank evaluation department, IEG, looked at what had happened to basic education (grades 1 to 9, in primary and junior secondary school) over this period.

Data and methodology

In 1988/89 Ghana Statistical Service (GSS) undertook the second round of the Ghana Living Standards Survey (GLSS 2). Half of the 170 areas surveyed around the country were chosen at random to have an additional education module, which administered math and English tests to all those aged 9-55 years with at least three years of schooling and surveyed schools in the enumeration areas. Working with both GSS and the Ministry of Education, Youth and Sport (MOEYS), IEG resurveyed these same 85 communities and their schools in 2003, applying the same survey instruments as previously. In the interests of comparability, the same questions were kept, although additional ones were added pertaining to school management, as were two whole new questionnaires – a teacher questionnaire for five teachers at each school and a local language test in addition to the math and English tests. The study thus had a possibly unique data set – not only could children's test scores be linked to both household and school characteristics, but this could be done in a panel of communities over a fifteen year period. The test scores are directly comparable since exactly the same tests were used in 2003 as had been applied fifteen years earlier.

There was no clearly defined 'project' for this study, rather support to the sub-sector through four large operations. The four projects had supported a range of activities, from rehabilitating school buildings to assisting in the formation of community-based school management committees. To identify the impact of these various activities a regression-based approach was adopted which analyzed the determinants of school attainment (years of schooling) and achievement (learning outcomes, i.e. test scores). For some of these determinants – notably books and buildings – the contribution of the World Bank to better learning outcomes could then be quantified. The methodology thus adopted a theory-based approach to identify the channels through which a diverse range of interventions were having their impact.

As discussed below, the qualitative context of the political economy of education reform in Ghana at the time proved to be a vital piece of the story.

Findings

The first major finding from the study was the factual. Contrary to official statistics, enrolments in basic education have been rising steadily over the period. This discrepancy was readily explained: in the official statistics, both the numerator and denominator were wrong. The numerator was wrong as it relied on the administrative data from the school census, which had incomplete coverage of the public sector and did not cover the rapidly growing private sector. A constant mark up was made to allow for private sector enrolments, but the IEG analysis showed that had gone up fourfold (from 5 to 20% of total enrolments) over the 15 years. The denominator was based on the 1984 census with an assumed rate of growth which turned out to be too high once the 2000 census became available, thus under-estimating enrolment growth.

More strikingly still, learning outcomes have improved markedly: 15 years ago nearly two-thirds (63 percent) of those who had completed grades 3-6 were, using the English test as a guide, illiterate. By 2003 this figure had fallen to 19 percent. The finding of improved learning outcomes flies in the face of qualitative data from many, though not all, 'key informant' interviews. But such key informants display a middle class bias which persists against the reforms which were essentially populist in nature.

Also striking are the improvements in school quality revealed by the school-level data: For example:

- In 1988, less than half of schools could use all their classrooms when it was raining, but in 2003 over two-thirds can do so.

- Fifteen years ago over two-thirds of primary schools reported occasional shortages of chalk, only one in 20 do so today, with 86 percent saying there is always enough.

- The percentage of primary schools having at least one English textbook per pupil has risen from 21 percent in 1988 to 72 percent today and for math books in Junior Secondary School (JSS) these figures are 13 and 71 percent, respectively.

School quality has improved across the country, in poor and non-poor communities alike. But there is a growing disparity within the public school sector. Increased reliance on community and district financing has meant that schools in relatively prosperous areas continue to enjoy better facilities than do those in less well off communities.

The IEG study argues that Ghana has been a case of a quality-led quantity expansion in basic education. The education system was in crisis in the seventies; school quality was declining and absolute enrolments falling. But by 2000, over 90 percent of Ghanaians aged 15 and above had attended school compared to 75 percent 20 years earlier. In addition, drop-out rates have fallen, so completion rates have risen: by 2003, 92 percent of those entering grade 1 complete Junior Secondary School (grade 9). Gender disparities have been virtually eliminated in basic enrolments. Primary enrolments have risen in both disadvantaged areas and amongst the lowest income groups. The differential between both the poorest areas and other parts of the country, and between enrolments of the poor and non-poor, have been narrowed but are still present.

Statistical analysis of the survey results showed the importance of building school infrastructure on enrolments. Building a school, and so reducing children's travel time, has a major impact on enrolments. While the majority of children live within 20 minutes of school,

some 20 percent do not and school building has increased enrolments among these groups. In one area surveyed, average travel time to the nearest school was cut from nearly an hour to less than 15 minutes with enrolments increasing from 10 to 80 percent. In two other areas average travel time was reduced by nearly 30 minutes and enrolments increased by over 20 percent. Rehabilitating classrooms so that they can be used when it is raining also positively affects enrolments. Complete rehabilitation can increase enrolments by as much as one third. Across the country as a whole, the changes in infrastructure quantity and quality have accounted for a 4 percent increase in enrolments between 1988 and 2003, about one third of the increase over that period. The World Bank has been the main source of finance for these improvements. Before the first World Bank program communities were responsible for building their own schools. The resulting structures collapsed after a few years. The Bank has financed 8,000 school pavilions around the country, providing more permanent structures for the school which can better withstand the weather.

Learning outcomes depend significantly on school quality, including textbook supply. Bank-financed textbook provision accounts for around one quarter of the observed improvement in test scores. But other major school-level determinants of achievement such as teaching methods and supervision of teachers by the head teacher and circuit supervisor have not been affected by the Bank's interventions. The Bank has not been heavily involved in teacher training and plans to extend in-service training have not been realized. Support to "hardware" has been shown to have made a substantial positive contribution to both attainment and achievement. But when satisfactory levels of inputs are reached — which is still far from the case for the many relatively deprived schools — future improvements could come from focusing on what happens in the classroom. However, the Bank's one main effort to change incentives — providing head teacher housing under the Primary School Development Project in return for the head teacher signing a contract on school management practices — was not a great success. Others, notably DFID and USAID, have made better progress in this direction but with limited coverage.

The policy context, meaning government commitment, was an important factor in making the Bank's contributions work. The government was committed to improving the quality of live in rural areas, through the provision of roads, electricity and schools, as a way of building a political base. Hence there was a desire to make it work. Party loyalists were placed in key positions to keep the reform on track, the army used to distribute textbooks in support of the new curriculum in the early 1990s to make sure they reached schools on time, and efforts made to post teachers to new schools and make sure that they received their pay on time. Teachers also benefited from the large civil service salary increase in the run up to the 1992 election.

Better education leads to better welfare outcomes. Existing studies on Ghana show how education reduces fertility and mortality. Analysis of IEG's survey data shows that education improves nutritional outcomes, with this effect being particularly strong for children of women living in poorer households. Regression analysis shows there is no economic return to primary and JSS education (i.e. average earnings are not higher to children who have attended primary and JSS compared to children who have not), but there is a return to cognitive achievement. Children who attain higher test scores as a result of attending school can expect to enjoy higher income; but children who learn little in school will not reap any economic benefit.

Some policy implications

The major policy finding from the study relates to the appropriate balance between hard and software in support for education. The latter is now stressed. But the study highlights the importance of hardware: books and buildings. It was also of course important that teachers were in their classrooms: government's own commitment (borne out of a desire to build political support in rural areas) helped ensure this happened.

In the many countries and regions in which educational facilities are inadequate then hardware provision is a necessary step in increasing enrolments and improving learning outcomes. The USAID project in Ghana encourages teachers to arrange children's desks in groups rather than rows – but many of the poorer schools don't have desks. In the words of one teacher, "I'd like to hang posters on my walls but I don't have posters. In fact, as you can see, I don't have any walls".

These same concerns underlie a second policy implication. Central government finances teacher's salaries and little else for basic education. Other resources come from donors, districts or the communities themselves. There is thus a real danger of poorer communities falling behind, as they lack both resources and the connections to access external resources. The reality of this finding was reinforced by both qualitative data – field trips to the best and worst performing schools in a single district in the same day – and the quantitative data, which show the poorer performance of children in these disadvantaged schools. Hence children of poorer communities are left behind and account for the remaining illiterate primary graduates which should be a pressing policy concern.

The study highlighted other areas of concern. First amongst these is low teacher morale, manifested through increased absenteeism. Second is the growing importance of the private sector, which now accounts for 20 percent of primary enrolments compared to 5 percent 15 years earlier. This is a sector which has had limited government involvement and none from the Bank.

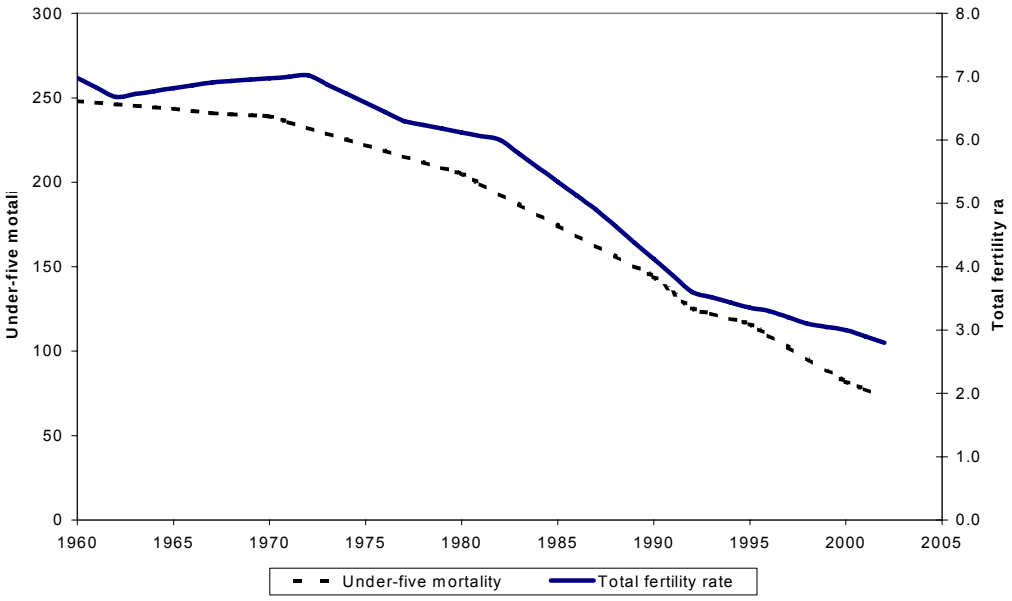
Annex IV: Meeting the Health MDGs in Bangladesh

Introduction

Bangladesh began the 1970s as a new country in a dire situation. The ravages of war and famine meant that the prospects for development appeared bleak. Social indicators were among the worst in the world. Women could expect to have on average seven children during their child-bearing years, but two of those would die before reaching their fifth birthday. Three-quarters of all children were malnourished.

Thirty years on the situation has changed drastically. The total fertility rate has fallen from seven to less than three, and under-five mortality from over 250 per 1,000 live births to around 80 by 2004 (Figure 1). The World Bank's (IEG's) reworking of the DHS data contested the widely-held view that fertility decline had reached a plateau in the 1990s. These rates of progress mean that Bangladesh is on track to meet the Millennium Development Goals. Malnutrition remains high but has begun to decline in the last decade. IEG (2005) examined the factors underlying this success.

Figure 1 Both fertility and under-five mortality have fallen



Methodology

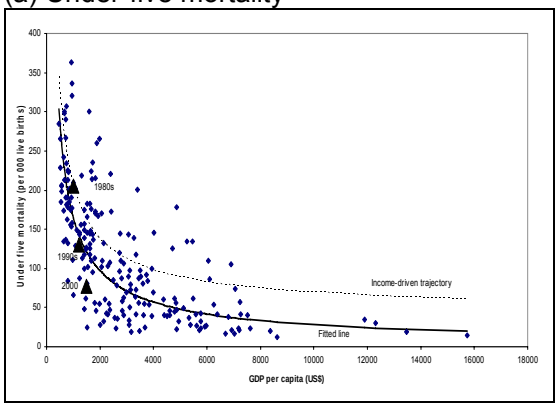
The IEG study utilized existing data sets. The analysis drew on both cross-country data, from a variety of sources, and national data mainly from the Demographic Health Surveys of 1992/93, 1996/97 and 1999/00. Multivariate analysis of the determinants of health and nutrition outcomes was carried out. This approached allowed the identification of interventions in a range of sectors which had affected health outcomes. While it was possible to carry out cost-effectiveness analysis, a full theory-based approach could not be applied because of the absence of process indicators.

Findings

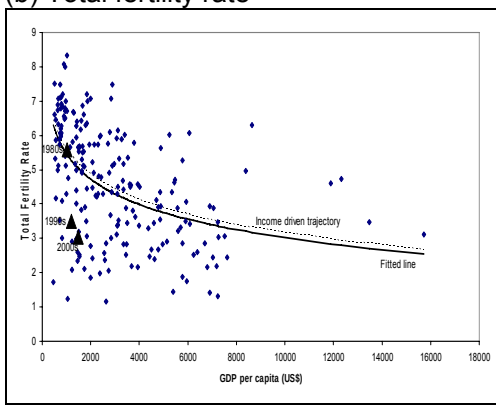
Economic growth is usually seen as a critical factor in reducing poverty in its various dimensions. Bangladesh is no exception to this point, and the country's respectable growth record has indeed played a part in the country's improved social outcomes. But it is not the whole story. Figure 2 shows under-five mortality and fertility plotted against income per capita for a cross-section of 78 countries at different points in time. Each data point represents the decade averages of income and the social outcome shown, using values from the 1970s to the current decade, so that there are up to four observations for each country. The solid line in each figure is the average relationship between income and the social outcome (that is "the fitted line"). In the 1980s Bangladesh (indicated by the triangular data points, each labeled by its decade) lay above the average for under-five mortality and fertility, meaning that those indicators were worse than should be expected for a country at its income level. If these indicators had improved following the internationally-established relationship with income then subsequent observations for Bangladesh would have laid along the dashed line. But in fact these later observations lie below the fitted line, showing that Bangladesh now does better than expected for a country at its income level. This finding suggests that there have been important, non-income-related, factors behind the improvement in mortality and fertility in Bangladesh.

Figure 2 Bangladesh's improvement in social outcomes is greater than can be explained by economic growth alone

(a) Under-five mortality



(b) Total fertility rate



Note: See discussion in text for explanation.

Table 1 Growth in GNP Per Capita Accounts for at Most One-Third of The Reduction in Mortality… and Less Than a Fifth of Lower Fertility

	1980 actual	**2000 actual**	**2000 income-based estimate**	**Percent reduction explained by income**
Under-five mortality	205.0	77.5	163.1	32.9
Total fertility rate	5.6	3.0	5.2	16.0

Source: calculated from data used for Figure 2.

The numbers behind these graphs provide an upper estimate of the extent to which growth in GDP per capita has contributed to improved social outcomes in Bangladesh (Table 1). For example, under-five mortality was 205 per 1,000 live births in the 1980s. Income growth alone would have reduced it to 163 by 2000, but by then the actual rate was 78. Hence at

most just under one-third of the improvement comes from higher average income. For fertility the share of income is even less, explaining at most 16 percent of the observed reduction. The question of what then explains the additional reduction was analyzed through multivariate analysis of both cross-country and household data. The results revealed the following regarding selected interventions:

- Immunization coverage was at less than 2 percent in the early 1980s, but grew in the latter part of the decade (largely with the support of UNICEF, but later also other donors including the World Bank) so that by 1990 close to half of all children were fully vaccinated in their first 12 months. Immunization has averted over 2 million child deaths in the last two decades, at a cost of between $100 and $300 per life saved.

- The World Bank financed the training of approximately 14,000 traditional birth attendants (TBAs) until the late nineties, at which point training TBAs was abandoned following a shift in international opinion toward a policy of all births being attended by Skilled Birth Attendants. However, the evidence presented in this report shows that training TBAs saved infant lives, at a cost of $220-800 per death averted.

- Female secondary schooling expanded rapidly in the 1990s, especially in rural areas partly as a result of the stipend paid to all female students in grades 6-10 in rural areas supported by Norwegian aid, the Asian Development Bank, the World Bank and government. Amongst the benefits of the increase in female secondary schooling are lower infant and child mortality, at a cost of $1,080-US$5,400 per death averted.

- Rural electrification, supported through three World Bank programs in the 1980s and 1990s, reduces mortality through income effects, improving health services, making water sterilization easier and improving access to health information, especially from TV. Taking these various channels into account means that children in households receiving electrification have an under-five mortality rate 25 per 1,000 lower than that of children in non-electrified households.

Policy implications

The IEG study had the following policy implications:

Publicly-provided services, with external support, were an efficacious and cost-effective means of improving health outcomes.

Interventions from several sectors improved health outcomes. But this multi-sectoral causation did not mean that interventions had to be delivered in a multi-sectoral manner.

Local evidence needs to be taken into account in making resource allocation decisions. The training of TBAs was abandoned in Bangladesh following international fashion, but local evidence shows it to have been effective in reducing infant mortality.

Draft 2
29 /11/07

NONIE Meeting
14 January, 2008
Washington DC

Testing Pawson's "Simple Principles for the Evaluation of Complex Programmes"[13]

For Discussion within Subgroup No.3

Nicolas Mathieu
Evaluation Department, EBRD

1. Introduction

1.1 Program theory

The core axiom of a Program theory is to make explicit the underlying assumptions about how an intervention is supposed to work and then to use this theory to guide evaluation. Programmes are seen as intervention chains, with one set of stakeholders providing resources to other sets of stakeholders, in the expectation (or 'theory') that behavioural change will follow. The success of the intervention is thus matter of the integrity of the sequence of programme theories and, in particular, how different stakeholders choose to respond to them.

1.2 EBRD Financial Sector Policies Evaluation (EvD 2007)

This Evaluation Special Study by the Evaluation Department of EBRD (EvD) is a review of the EBRD's Financial Sector Operations Policies (FSOPs) and performance in the financial sector in the Bank's countries of operations. The study has been carried out by the Bank's independent Evaluation Department. The review is also intended to assist the Bank in updating its policy for the financial sector. The principal objectives of the review are: (a) to trace the development of the Bank's financial sector operations policies through the policy documents adopted 1999; (b) to assess the Bank's performance in the financial sector: relevance[14], efficacy[15], efficiency[16] and impact at sector level[17]; and (c) drawing on findings

[13] R. Pawson (2005) 'Simple Principles for The Evaluation of Complex Programmes' in Killoran A, Kelly M, Swann C, Taylor L, Millward L, Ellis S (eds) *Evidence Based Public Health*, Oxford: Oxford University Press.
[14] Relevance is defined the relevance of each operation to the Bank policies stated in the program.
[15] Efficacy is defined here as the outcome of the program, i.e. fulfilment of program stated objectives.
[16] The proxy for efficiency at sector level is the average profitability of Bank operations in the sector.
[17] Transition impact in a sector evaluation is defined as the observable effect (through indicators and benchmarks) of the Bank's sector program (implemented with loans, investment, TA, policy dialogue) *on progress towards full market economy*, especially in: (1) structure and extent of markets, (2) market supporting institutions and (3) policies and market-based behaviour (See Table1).

of the review and other Bank material, to help identify challenges and opportunities for the future. In assessing the Bank's performance in the financial sector, the study focuses in particular on the 1999 financial sector policy and on the manner in which investment operations signed between late 1999 and 2006 responded to the priorities set out in the 1999 policy document[18].

2. Assumptions on how intervention is supposed to work

2.1 The EBRD Interventions in the Financial Sector

2.1.1. Population of projects. A total of 302 Financial Institution (FI) operations, comprising standalone projects and frameworks, were signed between the final quarter of 1999 and 2006. These do not include operations handled exclusively by the Group for Small Business. In addition there were a number of Technical Cooperation (TC) operations designed primarily with institution building in mind and a series of investment climate initiatives in the form of policy dialogue.

2.1.2. ***Sample for the evaluation***. The findings and conclusions of this study are based in large part on a detailed assessment of a sample of 101 signed operations and the manner in which they responded to the 1999 financial sector policy objectives. The sample, representing approximately one-third of operations signed in the seven-year period, comprises 40 operations evaluated independently by EvD, 31 operations for which bankers prepared self-assessment reports, but which were not evaluated by EvD, and a further 30 operations selected by the evaluation team to obtain a reasonable distribution of the total sample across countries of operations and project types.

2.2 The channels through which the program is to be implemented

> "The key to achieving a well functioning and stable financial system is to strengthen market mechanisms and initiatives and to support them with effective financial laws and regulations. The Bank's financial sector operations policy should thus focus on the overall objective of promoting **confidence** and **competition** in an independent financial system. This policy aims to translate these broad objectives into concrete criteria for the **selection and design of investment projects** in the financial sector and to identify priorities for **investment climate initiatives.**"[19]

2.3 The theory that behavioral change will follow

The implementation of the Policies, and the results of the Bank's various operations in the sector, should be evaluated with reference to the declared and approved overall objective of promoting confidence and competition in an independent financial system. The primary focus of the policy objectives enumerated in the document was **sustainability** (EvD emphasis) to

[18] It is worth recalling that the 1999 policy was prepared in the aftermath of the 1998 banking crisis in Russia. The policy document noted: "The Russian crisis and its repercussions throughout the region, in particular, have revealed the considerable risks in transition banking, arising from both macroeconomic instability and structural weaknesses within the banking sector itself. These risks are greatest in countries at earlier stages of transition."
[19] Financial Sector Operations Policy, BDS99-63 (Final), page 20 – emphases as in original.

be achieved by promoting confidence and competition through attention to the increased role of market processes and of government. High priority was given to strengthening the commercial orientation of state-owned financial institutions, supporting privatisation, facilitating foreign direct investment and promoting the consolidation of financial institutions through mergers and acquisitions. Government also had a role in developing legal, regulatory and supervisory frameworks that would permit financial institutions to operate in a market environment. The Bank would seek to help create new markets for financial services or extend their reach to new customers. The policy envisaged a strengthened corporate governance role for the Bank vis-à-vis its clients with attention also to the transfer of skills and technologies needed in a well functioning market-oriented financial system.

3. Pawson's first principle: locate key Program components

3.1 Principle

Evaluation should begin with a comprehensive scoping study mapping out the potential conjectures and influences that appear to shape the programme under investigation. One can envisage stage-one mapping as the hypothesis generator. It should alert the evaluator to the array of decisions that constitute a programme, as well as providing some initial deliberation upon of their intended and wayward outcomes.

3.2 Test

Table 1: Policy objectives linked to investment projects

Transition Categories	Transition Impact Indicators	FINANCIAL SECTOR POLICY OBJECTIVES RELATING TO INVESTMENT PROJECTS (1999 POLICY)
Structure and Extent of Markets	1. Competition	1.1 Add to the alternative sources of financing
		1.2 Promote efficiency and client orientation to increase competition
	2. Market Expansion	2.1 Extend financial sector coverage to new types of customer, especially private sector & SMEs
		2.2 Broaden range of providers of financial instruments (banks, non-banks and specialised institutions)
Market Supporting Institutions and Policies	3. Private Ownership	3.1 Strengthen commercial orientation of state-owned FIs through restructuring and privatisation
		3.2 Facilitate foreign direct investment
	4. Frameworks for Markets	4.1 Consolidation of financial institutions through mergers and liquidation
Market-Based Behaviour	5. Skills Transfer	5.1 Transfer skills and technology to local FIs
	6. Demonstration Effects	6.1 Broaden geographical coverage of the financial sector; support expansion of local institutions that have performed well
	7. New Standards for Business Conduct	7.1 Improve governance through participation in Boards
		7.2 Strengthen business practices of local FIs
		7.3 Improve transparency and disclosure standards

4. Pawson's second principle: prioritize among Program components

4.1 Principle

The general rule here is to concentrate on: i) those components of the programme theory which seem likely to have the most significant bearing on overall outcomes, and ii) those segments of programme theory about which least is known. Method-driven designs generally fall into the trap of trying to capture the whole, rather like a tailor sizing up the client's body. Theory-driven designs, by contrast, are light and strategic. It is better to draw out and test thoroughly a limited number of really key programme theories rather than achieve an approximate sketch of it all.

4.2 Test

It can be seen from the above Table 1 that each transition category and indicator was targeted by one or more policy objective. If the number of objectives is taken as a guide to the weight given to a transition category, then the table illustrates that 'market-based behaviour' was the

main target of the policy, followed by 'structure and extent of markets' and 'market supporting institutions and policies'. It seems appropriate that investment project objectives should target primarily issues and challenges that investment operations are capable of addressing. Therefore the policy objectives appear to be reasonably well targeted to address the Bank's transition goals. The policy objectives reflect the new emphasis on sustainability - **Market-Based Behaviour** to build confidence - together with a re-emphasis of financial deepening - **Structure and Extent of Markets** (competition and market expansion).

5. Pawson's third principle: evaluate Program Components by sub-sets

5.1 Principle

This principle is about when and where to locate evaluation effort in relation to a programme. The evaluation should take *on sub-sets* of programme theory. Evaluation should occur in ongoing portfolios rather than one-off projects. Suites of evaluations and reviews should track programme theories as and wherever they unfold.

5.2 Test

Table 2 : Transition Impact of sample investment projects[20]

TI Category/ TI Indicator	**No. of FSOP objectives**	**No. of projects addressing objectives**	**Average transition impact rating**
Structure and Extent of Markets	***4***	***96***	***2.32***
Competition	2	58	2.35
Market Expansion	2	86	2.31
Market-Supporting Institutions & Policies	***3***	***33***	***2.37***
Private Ownership	2	26	2.58
Frameworks for Markets	1	9	1.78
Market-Based Behaviour	***5***	***66***	***2.17***
Skill Transfer	1	29	2.10
Demonstration Effects	1	20	2.10
Improved Standards	3	55	2.21
Overall	***12***	***101***	***2.27***

Rating key: 1 = excellent; 2 = good; 3 = satisfactory; 4 = marginal; 5 = unsatisfactory; 6 = negative.

The Table 2 shows an overall average impact rating of close to *Good*. Interestingly, both the highest and lowest scores apply to the category "market-supporting institutions and policies". The analysis suggests that the "private ownership" indicator achieved least success while "frameworks for markets" (in practical terms, largely consolidation) scored highest.

[20] The first three columns in Table 2 replicate Table 1 and show (a) the standard TI categories and indicators against which the 1999 FSOP has been benchmarked; (b) the number of FSOP objectives addressing each TI category and indicator; and (c) the number of projects in the sample addressing the objectives. Column 4 shows the average rating assigned to each of the TI categories and indicators.

.

6. Pawson's fourth principle: Identify bottlenecks in Program network

6.1 Principle

"Theories-of-change" analysis perceives programmes as implementation chains and asks, 'what are the flows and blockages as we put a programme into action?' The basic strategy is to investigate how the implementation details sustain or hinder programme outputs. The main analytic effort is directed at configurations made up of selected segments of the implementation chains across a limited range of programme locations.

6.2 Test

Table 3: Transition Impact and Remaining Challenges 2005

Realised TI (i)	**Degree of Remaining Transition Challenges (ii)**		
	Small	**Medium**	**Large**
Highly significant Romania Serbia and Montenegro Slovak Republic	Slovak Republic	Romania	Serbia and Montenegro
Significant Bosnia & Herzegovina FYR Macedonia Georgia Kazakhstan Moldova Poland Slovenia	Kazakhstan Poland Slovenia	Bosnia & Herzegovina FYR Macedonia	Georgia Moldova
Moderate Armenia Azerbaijan Belarus Bulgaria Croatia Czech Republic Kyrgyz Republic Lithuania Russia Ukraine	Bulgaria Croatia Czech Republic Lithuania	Armenia Kyrgyz Republic Ukraine	Azerbaijan Belarus Russia
Minimal Albania Estonia Hungary Latvia Tajikistan Turkmenistan Uzbekistan	Estonia Hungary (N) Latvia	Albania	Tajikistan Turkmenistan Uzbekistan

(Sources: (i) EBRD Transition Impact Retrospective 2, Companion Paper CS/FO 05-17;
(ii) Assessment of Transition Challenges CS/FO 05-10).

While there has been moderate impact in Azerbaijan, Belarus and Russia, there remain large challenges in these countries, largely due to the work still to be done on "**market-supporting institutions and policies**", in particular effective regulation and supervision. Serbia and

Montenegro, Georgia and Moldova seem to be extreme cases where the Bank records significant or highly significant impact, but large challenges remain[21].

7. Pawson's fifth Principle: Provide feedback on conceptual framework

7.1 Principle

What the theory-driven approach initiates is a process of *'thinking though'* the tortuous pathways along which a successful programme has to travel. What would be described are the main series of decision points through which an initiative has proceeded and the findings would be put to use in alerting stakeholders to the caveats and considerations that should inform those decisions. The most durable and practical recommendations that evaluators can offer come from research that begins with theory and ends with a refined theory.

7.2 Test

7.2.1 More focused policy priorities

At the commencement of the present financial sector policy assessment, the evaluation team extracted the investment project and investment climate related objectives from the 1999 policy document. While the document was sufficiently clear to allow the priorities to be identified, their presentation in the policy document might be regarded as somewhat diffuse which could lead to a loss of focus in implementation over time. Future statements of policy should relate strategic priorities clearly to transition challenges and transition gaps. An indication of strategic targets should be given to enable progress to be assessed periodically and policy refinements to be adopted in the light of experience.

7.2.2. Re-emphasise the importance of market-supporting policies

The 1999 FSOP stressed the importance of encouraging the development of sustainable financial systems though increasing the role of market processes. In the evaluation team's view the 1999 FSOP targeted principally 'market supporting institutions and policies', followed equally by 'structure and extent of markets' and 'market-based behaviour'. A conclusion of this Study is that, from the sectoral perspective, the implementation of the policy has focused more on the latter two transition categories, certainly as concerns the objectives of investment operations. This translates into a focus on competition, market expansion and institution building which has contributed strongly to the ongoing transition process in those countries which already have sufficient market supporting institutions and policies in place. A challenge facing the Bank is to design investment operations in such a way as to stimulate and encourage the development of sound legal, regulatory and supervisory frameworks in countries where these are still lacking.

[21] Using "country performance comparators" i.e. "countries where the Bank did not intervene" for impact evaluation purposes is not feasible in this case, since EBRD intervened in all the countries where it has a transition support mandate. What can be done on the basis of the classification of the Table 1, however, is to (1) compare the sector impact of countries with similar initial level of transition and similar intensity of Bank support, and (2) link the level of " remaining challenges" to (a) the type of support that the countries have received and (b) the country's business environment. For example: Romania and Slovak Republic, Bulgaria and Belarus, Albania and Uzbekistan.

8. Conclusion: relevance of Pawson's evaluation principles to applied work

The principles appear useful to link program, objectives to various channels of implementation and to locate specific outcomes and bottlenecks. They also seem to be appropriate to see the extent to which a program had balanced components to start with and why the same program may not carry its expected impact due to unbalanced applications. Some program channels can overused while others are neglected. Overall this approach helps fill a gap in the current evaluation literature which does not pay enough attention to the interaction between program processes and program outcomes.

9. Next Step: towards a quantification of impact of program interventions

While experimental and quasi-experimental methods are well developed to measure the impact of investment projects, they are an early stage when it comes to measure of program interventions at sector level. This is because programs processes appear to be much more complex than project processes[22]. Programs involve policy instruments, each one carrying various channels through which it can be activated, and several targets, each one of these targets having its own set of beneficiaries. We are in presence of policies applied to networks of *potential* beneficiaries where each one can be associated or separated from the benefits of the program, and when associated, receive the benefit with various intensities.

The above problem of sector impact maximization could take the form a dynamic program optimisation applied to a "social network". A social network of program beneficiaries can be modelled as a set of entities or actors with binary relations between any two of them. Each actor is regarded as a vector of time with varying deterministic and/or random variables. The vector may consist of a dynamic set of statistical information, attributes about other actors, and a set of personal values, attributes, and preferences. The binary relation between any two actors depends on the maximum payoff of an appropriately constructed nonlinear programming problem involving the attributes and values of the actors in the social group.

This branch of non linear programming could eventually be used in studying social impacts and improving support to policy and program interventions[23]. Fortunately, the statistics literature on program optimization for social networks has recently emerged from the more classical approaches of maximization of welfare functions through optimal processes applied so far to macro economic policies.

[22] Except when a sector program consists in an aggregation of investment projects.

[23] See H.T. Banks, *A stochastic Dynamical model for Social Networks*, . and Hong, Chung-Chien , *Nonlinear Programming and Optimal Control Approach To the study of Social Network*, Center for Research in Scientific Computation , North Carolina State university in 3rd International Conference on Neural, Parallel and Scientific Computations, August 9-12, 2006, Moore House College, IFNA, Atlanta, GA, USA.