

Research Paper No. 2007/52

Monitoring and Evaluation Reform under Changing Aid Modalities

Seeking the Middle Ground
in Aid-Dependent Low-Income Countries

Nathalie Holvoet and Robrecht Renard*

September 2007

Abstract

This paper grew out of our bewilderment with the insouciance with which some in the donor community seem ready to abandon accounting for the use of aid. If one listens to the rhetoric surrounding the new approach to aid, one gets the impression that most of the crucial accounting tasks must be swiftly abandoned by donors and left to recipient governments. This paper does not question the underlying rationale for shifting towards recipient-led priority setting and control over implementation of aid resources, but argues that donors cannot let themselves off the hook so easily with respect to the accountability part of the equation. We argue that in most low-income countries such trust in recipient systems may be dubbed as over-alignment, and that it is neither necessary nor useful. Our argument is however not that old style donor-managed monitoring and evaluation is the only or the best solution. For we are equally puzzled by the stubbornness with which some other donors stick to their old monitoring and evaluation (M&E) in ways that contradict the new insights in aid effectiveness and hamper the emergence of

.../.

Keywords: aid, modalities, reform, accountability, feedback, alignment, diagnosis of monitoring and evaluation, low-income recipient countries

JEL classification: F53, O19

Copyright © UNU-WIDER 2007

* Institute of Development Policy and Management, University of Antwerp, Antwerp, emails: nathalie.holvoet@ua.ac.be (N. Holvoet) and robrecht.renard@au.ac.be (R. Renard)

This study is a revised version of the paper presented at the 16-17 June 2006 WIDER development conference on 'Aid: Principles, Policies, and Performance', directed by George Mavrotas.

UNU-WIDER gratefully acknowledges the financial contribution to the conference by the Finnish Ministry for Foreign Affairs.

UNU-WIDER also acknowledges the financial contributions to the 2006-07 research programme by the governments of Australia (AusAID), Denmark (Royal Ministry of Foreign Affairs), Norway (Royal Ministry of Foreign Affairs), Sweden (Swedish International Development Cooperation Agency—Sida), and the United Kingdom (Department for International Development—DFID).

national M&E systems. Why are positions so polarized and why is hardly anyone arguing in favour of intermediate positions? This is what this paper sets out to do: we argue against a radical and rapid implementation of the new rhetoric in low-income countries, but also against a continuation of present accountability practices. Donors have a large and lasting responsibility in accounting for the use of aid funds, both towards the taxpayers in donor countries and towards the targeted beneficiaries in the at best pseudo-democratic and poorly governed low-income recipient countries. They should find new ways to remain firmly involved in M&E, ways that allow, at the same time, embryonic national M&E systems in low-income recipient countries to grow and flourish.

Acronyms

APR	annual progress report
CDF	comprehensive development framework
CSO	civil society actors
ECD	evaluation capacity development
H&A	harmonization and alignment, conference 2003 Rome
HIPC	heavily indebted poor countries
IEO	Independent Evaluation Office of the IMF
JRM	joint sector review missions
JSA	joint staff assessments
M&E	monitoring and evaluation
MDG	Millennium Development Goals
PBAs	programme-based approaches
PEFA	public expenditure and financial accountability
PEM	public expenditure management
PFM	public finance management
PRSPs	poverty reduction strategy papers
PSIA	poverty and social impact analysis
RBM	results-based management
SSA	Sub-Saharan Africa
SWAPS	sector-wide approaches

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

www.wider.unu.edu

publications@wider.unu.edu

UNU World Institute for Development Economics Research (UNU-WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Typescript prepared by Liisa Roponen at UNU-WIDER

The views expressed in this publication are those of the author(s). Publication does not imply endorsement by the Institute or the United Nations University, nor by the programme/project sponsors, of any of the views expressed.

Introduction

The shift in aid modalities characterized by a move towards more programme-based approaches (PBAs),¹ demands considerable behavioural changes of both recipients and donors. In the era of poverty reduction strategy papers (PRSP)² and sector wide approaches (SWAPS), donors are moving away from micro-management and full control over all stages of earmarked aid interventions, towards arm-length control and supervision over recipient policies and implementation procedures. In practice, this involves more emphasis on policy dialogue; when possible full alignment with a recipient country's national system of planning, budgeting, implementation, monitoring and evaluation, and where this is not 'yet' deemed 'appropriate', harmonization with other donors and minimizing parallel structures for aid delivery; and investment in recipient-driven capacity-building. It is expected that recipients elaborate sound policies; strengthen (results-oriented) institutions for planning, implementation, monitoring and evaluation; and take leadership in the identification of weaknesses in policy and institutional apparatus and elaboration of plans for improvement and capacity-building.

This paper focuses on the consequences in the area of monitoring and evaluation (M&E). The salient features of the new approach—such as the decrease in donor earmarking, the increase in recipient responsibility for implementation, and the imposition by donors of a string of new process conditionalities that constrain the latitude of the recipients in their use of aid monies—all point to the need for a restyling and upgrading of M&E tasks. As explained in section 2, national M&E systems need be strengthened in their twin functions of accountability and lessons-learning, and likewise, the role of donors in M&E needs to be recast. With time, as recipient M&E systems become stronger, their output should be able to satisfy a major part of the accountability needs of donors, and this should provide a powerful impetus to wind down parallel donor M&E activities in a country.

Notwithstanding the fact that the new approach to aid has such a well-recognized incidence on M&E, section 3 claims that until recently, M&E issues have not figured highly on the recipients' reform agendas. Section 4 looks at progress on the donor side, and the picture that emerges is not much better. Effective realizations on the ground in general have been disappointing. Scant attention has been paid, for instance, by donors to M&E in assessing country-produced PRSP documents, which excel in 'indicatorism'

¹ The OECD-DAC (2005a: 37) document uses the notion of 'programme-based approaches' to refer to 'a way of engaging in development cooperation based on the principles of co-ordinated support for a locally owned programme of development'. The approach includes four key elements: (i) leadership by the host country or organization; (ii) a single programme budget framework; (iii) donor coordination and harmonization of procedures; and (iv) efforts to increase the use of local procedures over time with regard to programme design and implementation, financial management, monitoring and evaluation.

² Literally the PRSP is a policy document that is produced by the recipient country and replaces similar documents such as the policy framework paper that used to be imposed by the IMF and the World Bank on low-income aid-dependent countries for the purpose of adjustment lending. More fundamentally, it constitutes the linchpin of a new comprehensive aid approach that concentrates on four basic principles: (i) long-term and holistic vision, (ii) country ownership, (iii) results orientation and (iv) country-led partnership. Nowadays, donors increasingly use the PRSP as the framework for their aid activities, accompanying the shift from projects to sectoral and general budget support.

and in general remain obscure regarding the overall M&E policy, its institutional translation or the roles and responsibilities of the various parties involved. In recent years, several surveys and independent reviews have been conducted that confirm that progress towards results-orientation, and particularly a reinforced M&E, has been disappointing. The World Bank's 2004 PRSP evaluation study includes the results of a survey among almost 800 stakeholders in ten PRSP countries, which reveals that out of 39 questions the one on M&E received the most negative response (World Bank 2004: 66). In fact, to the question 'An effective structure to monitor and evaluate results has been established', 41 per cent of the respondents 'disagreed', or 'disagreed completely', whereas a further 21 per cent answered 'don't know or unsure'.

If attention to the effective elaboration of recipient M&E systems has been meagre, the donor part of the reform agenda has not been squarely addressed. While one could argue that PRSP documents are not the primary source of information on donor behaviour, one would expect at least that the involvement of donors in M&E activities would be acknowledged and the need for reforms explicitly addressed. Yet this is not the case. Similarly, issues of harmonization and alignment are hardly touched upon, notwithstanding the fact that the integration of parallel donor-driven project M&E systems is highly relevant for the elaboration and strengthening of national M&E. Surveys on donors' harmonization and alignment have in the meantime confirmed that this silence is testimony to the fact that not much is being done on the ground. Among others, the 2003 World Bank evaluation of the comprehensive development framework (CDF) highlights that M&E is one of the areas where donors have made least progress in harmonizing practices and where consequently the burdens on recipient government are enormous (World Bank 2003a).

Section 4 concludes that donors seem caught in a chicken-and-egg dilemma: most still overwhelmingly rely on their own institutional apparatus for aid delivery, are beginning to harmonize only slowly with other donors and hardly align with national apparatus they rightly consider weak. However, adherence to one's own parallel structures at the same time burdens national systems.

Recently, international pressure for harmonization and alignment has moved these topics up on the agenda: the 2003 Harmonization and Alignment (H&A) Conference in Rome; the Marrakech Memorandum as well as the 2005 *Paris Declaration on Aid Effectiveness* testify to this.³ Conferences and peer pressure seem to pay off, swaying some of the laggards among aid agencies. Belgium, a small donor with which the authors have extensive personal experience, may well be illustrative of what is going on in some other agencies. Over a period of five years, the mood has moved from a stubborn unwillingness by a vast majority of staff to even discuss the negative effects of old type project aid, through to a period of almost 'religious war' among irreconcilable factions (the wording was suggested to one of the authors by a senior staff describing the mood in another small European donor agency going through a similar process) to the victorious adoption of a very simplistic 'alignment' mantra by a growing number of leading aid officials, to the point of denying the considerable political, institutional and technical weaknesses of partner countries. As argued in section 4, this tendency may possibly grow stronger if future accountability to parliament and public opinion would be reduced to showing how much effort is spent on adopting the new H&A agenda

³ See www.aidharmonization.org.

rather than a thorough argumentation on how it will lead to sustainable poverty reduction on the ground.

With the limitations of the old-style isolated project approach now widely accepted, and the importance of harmonization and alignment generally acknowledged, we see the danger of some donors going overboard on the other side. Donor bureaucracies may well be tempted by a version of M&E alignment that is shortsighted and capable of undermining the new aid paradigm. The diabolizing of the project aid paradigm has partly been achieved by making a caricature of what are genuine constraints and problems facing most of the recipient systems in low-income countries. This may well explain to some extent the donor's neglect of recipients' M&E systems. Alignment to embryonic, and in many respects dysfunctional, national M&E systems in low-income countries, believing that these will somehow live up to the complex demands set by programme-based approaches may eventually be as counterproductive as bypassing these arrangements and keeping uncoordinated donor systems intact. Exemplary of the donor's underestimation of the huge challenges posed is the paucity of systematic diagnostic tools to assess recipient M&E systems. Such assessments should look at both supply- and demand-side components of M&E products, identify strengths and weaknesses, and assess alignment risks and ways of counterbalancing those risks in the short and long run. Recently, donors have been engaging timidly in fragmentary M&E assessments, but these are mainly limited towards data quality diagnosis.⁴ This is in stark contrast with the diagnostic 'bombardment' recipients suffered in the area of public expenditure management (PEM).⁵ This imbalance leads to an undue focus on financial inputs, at the expense of monitoring the overall causal chain and of evaluating cause-effect linkages, as well as to a bias in favour of expedient technical solutions that do not address the underlying institutional problems. All this, to our mind, is symptomatic of the lack of donor ambition and resolve in the field of M&E (see also Bamberger 1989, 2000; Picciotto 2003 and World Bank 2002). We do not mean to belittle the recent emphasis on poverty and monitoring of the Millennium Development Goals (MDG), which shifts the attention from inputs to the final outcomes/impact at the other extreme of the chain, or the growing donor engagement in poverty and social impact analysis (PSIA). But it is often done in a haphazard, disintegrated manner, and a coherent overall approach to M&E is mostly lacking. We believe that several of the weaknesses currently identified with programme-based approaches can, in fact, be traced back to M&E⁶ related issues:⁷ the low quality of the annual progress reports, the

⁴ To assess statistical capacity and to engage in capacity-building, several (diagnostic) frameworks exist. Bedi et al. (2006) describe the data-quality assessment framework (DOAF), the general data dissemination system (GDDS), the guide to designing a national strategy for the development of statistics (NSDS) and the statistical capacity-building indicators (SCBI). The WB Country Statistical Information Database contains information for 143 developing countries on some of the above-mentioned indicators and frameworks (see www.worldbank.org/data/countrydata/aboutcsidb.html and IMF-WB 2005).

⁵ The 2004 SPA *Budget Support Alignment Survey* indicated that public finance management (PFM) diagnostic exercises were an important area of uncoordinated donor review that imposed large burdens on recipient countries. See for more details SPA (2004: 15-18). In order to cope with this the public expenditure and financial accountability (PEFA) initiative has been initiated.

⁶ This is particularly true if you adopt an expanded definition of M&E. Boyle, Lemairre and Rist. (1999: 4-5) refer in this respect to Rist (1990: 4-5) who notes that 'the development of programme evaluation, both in terms of its methodologies as well as the kinds of questions it could address, has resulted in a clear expansion of what comes under its umbrella. ... But the most recent thinking

inconclusive reviews of national or sector programme-based approaches, as well as the conclusion that in many countries an action-oriented approach is lacking may all be linked to the absence of an underlying programme theory. As will be highlighted in section 5, applying the general insights from M&E theory can significantly contribute to understanding the observed weaknesses and point towards possible remedying interventions that would stimulate the implementation of genuine programme-based approaches and add to their sustainability.

Moving beyond the false dichotomy between a principled handing over of M&E responsibilities to the recipient country without any comprehensive prior assessment of the capacity of national M&E systems to take on such tasks on the one hand, and a continuation of non-harmonized and non-aligned donor M&E practices on the other hand, section 5 further proposes a pragmatic twin-track process approach to get out the present deadlock. Balancing between the need to align as much as possible to the recipient's M&E while filling existing M&E gaps and satisfying accountability needs towards citizens both in donor and recipient countries, we suggest an approach that focuses on strengthening both the demand and supply sides of recipient M&E, combined with complementary donor-steered M&E.

2 The M&E reform agenda

Programme-based approaches, compared to traditional projects, reduce the level of earmarking and intentionally blend donor resources with those of the recipient. This leaves recipients more in charge, and their responsibility in implementing policies is thereby enhanced. Donors, through process conditionalities, request a convincing national planning-budgeting-implementation cycle in return. They however do not believe in the top-down, blueprint planning that was practiced in the 1960s and 1970s in many low-income countries. Instead, they push for a policy cycle that relies on continued feedback from monitoring and evaluating progress. In doing so, they have also their own accountability needs in mind. They are fully aware that the final results that will hopefully validate their trust in recipient governments can only be gauged through proper evaluation, and that much of the data for this can only be provided by recipients. A well-functioning national M&E system that devotes considerable attention to the twin functions of feedback (lesson learning) and accountability is thus a linchpin of the new aid paradigm. That is why M&E should be high on donors' and recipients' lists of priorities. Conversely, without convincing recipient M&E, the willingness of the more reluctant donors to fully embrace the new approach will remain in doubt, as they will not be able to assess how their aid resources are being handled in the long lead time between the transfer of aid resources and outcomes/impact, and as it will be impossible to establish ex post what the impact has been of their aid, or why eventually things went wrong.

suggests that programme evaluation can encompass the various stages of life cycles of a programme or policy—from conception through execution through impact’.

⁷ Recently, this been acknowledged in the IMF-WB 2005 PRSP evaluation study that stresses the need for making M&E a central issue in PRSP as this may lead to more action-oriented PRSP.

The construction of a convincing PRSP as a prior condition that low-income countries must meet before they receive enhanced HIPC debt relief/budget support illustrates this donor concern. Significantly, this conditionality also illustrates the rediscovery by donors of the importance of good macro planning. What donors have in mind when they think about a good PRSP—and they make sure that this is clearly spelled out to partner countries—is a policy cycle approach towards poverty reduction whereby diagnosis, priority and strategy selection, resource allocation and budgeting, implementation, monitoring and evaluation are all logically linked. A lot of emphasis is put, for instance, on ‘diagnosis’ as the basis for selection of priorities and on the ground this has led to an effective upsurge in data collection, especially on poverty. Throughout every PRSP cycle, lasting typically 3-4 years, and even more so when moving from one PRSP to a new one, priorities and strategies are expected to be adjusted on the basis of information from monitoring and evaluation. At the same time, and with the same M&E effort, accounting needs must be addressed, both within the recipient country, and towards the donor community. This explains the importance of the yearly annual progress report (APR), whose objective is to summarize progress on targets identified, analyse underlying reasons for lack of achievement, and adjust targets where deemed necessary.

One of the major principles of new aid instruments is to shift attention away from inputs to results (defined in terms of outcomes and impact) and to the link between the two. If resource allocation and planning are to be done increasingly on the basis of results generated, then one needs information about results and how they derive from, or fail to derive from, inputs. Results-based management (RBM) thus generates a demand for M&E. In order to ensure the reliability of data and their analysis, M&E services must be sufficiently autonomous from those that are responsible for implementation.

Programme-based approaches, and the PRSP which serves as an overall framework, also foster participation of a broad range of stakeholders at the macro-level, basically because this is what donors want. The idea is that non-state stakeholders, and in particular those that directly represent beneficiaries and users of public services, can play key roles in producing ‘independent’ assessments of the implementation and impact of service delivery and policy processes. As local organizations and associations they have ‘grassroots’ contacts and are therefore well placed to channel and represent the voices of the beneficiaries, and to supply information for the M&E system. The implicit assumption is that such civil society actors are sufficiently ‘representative’ and sufficiently close to the poor to act on their behalf. These non-state actors will exert pressure for better and timely information, and strengthen the demand side of M&E.

That most donors now recognize that traditional project aid is an inappropriate aid delivery mechanism to aid-dependent countries has also influenced the M&E reform agenda. Several studies (see e.g., Eriksson 2001) have highlighted the havoc caused by donors imposing their separate and uncoordinated rules for aid management, and how this has severely undermined the institutional capacity of the aid-dependent governments to plan, budget and implement their own development. Concerted aid to government, respecting the latter’s own policy priorities, using national budgeting and allocation mechanisms, national implementation, supervision and control systems is the answer of the new approach to aid. In the field of M&E this means winding down parallel donor systems and replacing these by national M&E systems, whose reorganization and strengthening are the focus of yet another set of process conditionalities.

In practice, the reform agenda, including the one in M&E, has been shaped through various international conferences and related documents, such as the First International Roundtable on Managing for Results (2002); the Second International Roundtable on Managing for Results (2004),⁸ the Marrakech Memorandum (featuring five core principles); the Paris Harmonization, Alignment and Results High Level Forum (2005: see the Paris Declaration on Aid Effectiveness including twelve indicators for monitoring the donor and recipient reform agendas).

In summary, the expectations are that:

partner countries endeavour to establish results-oriented reporting and assessment frameworks that monitor progress against key dimensions of the national and sector development strategies and that these frameworks should track a manageable number of indicators for which data are cost-effectively available (Paris Declaration: indicator 11).

Additionally, partner countries should:

reinforce participatory approaches by systematically involving a broad range of development partners when formulating and assessing progress in implementing national development strategies (Paris Declaration: 9).

It is expected that donors ‘work with partner countries to rely, as far as possible, on partner countries’ results-oriented reporting and monitoring frameworks’ and that they:

harmonize their monitoring and reporting requirements, and, until they can rely more extensively on partner countries’ statistical, monitoring and evaluation systems, [work] with partner countries to the maximum extent possible on joint formats for periodic reporting (ibid.: 8).

From both partners it is expected that they ‘work together in a participatory approach to strengthen country capacities and demand for results based management’ (ibid.: 8).

The DAC guidelines (OECD-DAC 2003) make the M&E reform agendas of the donors operational by identifying good-practice indicators for reporting and monitoring, broken down according to the following five guiding principles: relying and building on country systems; coordinating reporting and monitoring systems; simplifying reporting and monitoring systems; making information more transparent; and rationalizing review missions.

3 Limited and unequal progress on the recipient side

If the M&E reform agenda is so crucial for the success and sustainability of the new aid paradigm, what can we learn from the actual practice in recipient countries? The

⁸ See www.mfdr.org/ for more information on Managing for Development Results, and the different Roundtables.

analysis in this section⁹ draws on a review of academic literature and practitioners' documents¹⁰ as well as our own mid-2004 desk study assessing on the basis of PRSP and APR the quality of recipient M&E systems in eleven Sub-Saharan Africa countries.¹¹ We also draw on the insights from our own fieldwork, with respect in particular to a field study in Ethiopia in July 2005 which completely focused on this issue. In general, we were puzzled by the relative lack of attention to M&E issues, both in country-produced PRSP documents and in joint staff assessments (JSA)¹² by IMF and WB staffs. For instance, most of the time, JSA made only a few bland general comments, of the variety: 'some progress has been made but much remains to be done'.

In what follows we will structure our discussion along the following headings: policy, institutional and organizational issues; methodology; participation of non-state actors and quality and capacity.

3.1 Policy, institutional and organizational issues

Independent reviews (see footnote 10 for an overview) as well as our own desk study underscore that there is a very fragmentary approach towards M&E, and that the focus is overwhelmingly on technical and methodological issues, to the detriment of the overall policy and the institutional and organizational set-up. Of the eleven countries reviewed in the desk study, most had some elements of an M&E plan, heavily skewed towards data collection issues, but hardly any country had a comprehensive 'grand design' or provided details about the overarching institutional structure.

One of the most glaring weaknesses in the M&E policy and design of recipient countries is the 'conflation' of monitoring and evaluation. The fact that they are often referred to as twin concepts is fine, but not when both functions are confused with each

⁹ We would like to stress that the results of the analysis hold only for low-income recipient countries, and in particular for the eleven SSA countries of our own desk study (see footnote 11). While some of the weaknesses (such as the defective linkage among M&E, budgeting and policy-making) pointed at in the analysis also hold for middle (and even high)-income countries, M&E systems of the latter are in many cases much more advanced. See, for instance May et al. (2006) for an overview of M&E systems in Chile, Colombia, Mexico, Brazil and Peru as well as several country studies on the website of the Evaluation Capacity Development group within the Independent Evaluation Group (see www.worldbank.org/ieg.ecd).

¹⁰ See, among others, Bedi et al. (2006); Booth and Lucas (2002); Evans and Ngalwea (2001); Evans and van Diesen (2002); GTZ/BMZ (2004); Hauge (2001, 2003); IMF and IDA (2003); IMF (2004); IMF and World Bank (2005); Lucas et al. (2004); Prenushi, Rubio and Subbarao (2003); McGee and Norton (2000); ODI (2003); Robb (1999); World Bank (2003a, 2003b, 2004, 2005).

¹¹ The desk study assesses the quality of PRSP M&E systems in eleven SSA countries: Burkina Faso, Ethiopia, Ghana, Malawi, Mali, Mauritania, Mozambique, Niger, Tanzania, Uganda and Zambia. The elaborated assessment framework (in the Appendix) captures issues of policy, methodology, organization, capacity, participation and quality. See Holvoet and Renard (2005, 2007) for a more detailed discussion of the methodology and the findings.

¹² JSAs are prepared by IMF and WB staff members and are submitted together with a country's PRSP to the Board of Directors. Initially, the JSAs had both a feedback and a signalling function: it highlighted the strengths and weaknesses of a country's PRSP and how it might be improved while it also indicated whether the PRSP was a sound basis for concessional assistance from the WB and the IMF. Since late 2004 the signalling function has been dropped and JSAs have been renamed as joint staff advisory notes. For more information about JSAs, see ODI (2004).

other, or one is subsumed in the other. Frequently the impression given is that monitoring somehow also constitutes evaluation. Evaluation, if mentioned as a separate issue at all, seems almost an afterthought. We acknowledge that in iterative planning, monitoring is of the utmost importance, and even that more energy is devoted to it than to evaluation, but there remains a persistent need for the latter. One of the consequences of this leaning towards monitoring at the expense of evaluation is that the focus is much more on pure stocktaking of performance ('were the targets met?') than on probing into underlying reasons for (non)performance ('why were the targets not met?') or on questioning the relevance and usefulness of some of the targets themselves and/or of specific interventions designed to arrive at them. Particularly neglected is the role 'programme theory evaluation' or 'theory-based evaluation' could play in the numerous cases where the causal chain behind an intervention is seemingly absent or at least not made explicit in the PRSP documents. Extracting, on the basis of relevant stakeholders' views, the underlying programme theory may help identifying the causes for (non)performance, differentiating between failures in implementation and underlying programme theory (on this, see, e.g., Kusek and Rist 2004; Rossi, Freeman and Lipsey 2004). Given that the role of evaluation theory is downplayed to such an extent, it comes as no surprise that annual progress and sectoral progress reports, the most important outputs of national and sectoral M&E systems, are often sloppy at best. In our view, donors have a major responsibility in this neglect of evaluation. They are generally more concerned with performance management and input monitoring than with higher-order policy-relevant questions. We find some support for our view in the *2002 Annual Report on Evaluation Capacity Development*: 'PRSPs and Bank/Fund JSA often use the term M&E as shorthand for financial tracking and national development indicators. And the recent Bank/Fund review of the PRSP approach also interpreted M&E as being synonymous with monitoring' (World Bank 2002: 20).

The neglect of evaluation also implies that there is hardly any attention to the linkages between monitoring and evaluation. The disintegrated way in which PSIA is currently dealt with is an illustration. While one may applaud the increasing appeal of PSIA, it is usually conceived outside the national M&E system, mostly performed by donors, with hardly any participation of national stakeholders. Information from national monitoring does not seem to feed into PSIA, neither are there guarantees that information from PSIA feeds into policymaking.

Somewhat related to this is the absence of a thorough discussion in national M&E plans of the different functions of M&E and basic guiding principles, and in particular the possible trade-offs among these. In fact, hardly any document indicates how M&E will fulfil both its basic but potentially conflicting functions of accountability and feedback and how related principles of impartiality, independence and utility will be guaranteed. Institutional arrangements do not seem to be driven by these questions of how to keep a balance between such functions and needs. For instance, it is quite difficult to get a clear perception of how independent M&E units, at either national or sector level, are from senior management and government ministers. Admittedly, learning, feedback and integration of M&E findings into management and policy call for close linkages and by the same token militate against too much independence of M&E. But on the other hand, the accountability function cries out for sufficient independence. What is worrisome is that such trade-offs between independence and feedback are not explicitly addressed. Issues of independence and credibility of M&E are sensitive issues in all countries, and particularly so in low-income recipient countries that have almost no tradition in transparency in public affairs. Independent M&E may disclose not only positive news,

but also negative findings and in particular unveil politically sensitive information about corruption, economic rent-seeking by the political elite, poor management by senior political appointees, etcetera. Most of the PRSP also remain remarkably vague about the location of the oversight function in M&E. In a similar vein, Bamberger (1991), in a case study of national M&E systems in South Asia, discovers that because of severe competition among various central ministries over the control of M&E, no institutional arrangement remained in place for longer than three years. The major reason for the constant changes was the fact that the agencies in charge of M&E were perceived to be becoming far too powerful and threatening.

What is also striking is the lack of clarity about roles and responsibilities of the actors involved in data collection, analysis and feedback, including statistical agencies, line ministries, decentralized levels and central ministries. In most countries, sector monitoring systems existed before the PRSP, and in a lot of cases they have been strengthened in the context of SWAPs. While one could logically conceive these sectoral systems as important pillars of a national coordinated M&E system, obscurity surrounding roles and responsibilities, and the absence of even rudimentary coordination and alignment mechanisms can easily lead to institutional competition and to dysfunction. In fact, one may consider the balance between independence and quality of sectoral and decentralized programmes and M&E on the one hand and their degree of integration with the national level crucial issues for the sustainability of the new aid paradigm. Obviously, moving towards results-based management and budgeting can only work where there is a minimal degree of horizontal and vertical integration. This involves line ministries and decentralized levels that supply central ministries with the necessary information on intermediate and final outcomes, and central ministries that provide line ministries and decentralized levels with realistic projections on future budgets. At the same time, however, too much focus on integration and centralization may put undue stress on sectoral and decentralized M&E systems if it means that they are forced to emphasize functions of accountability and feedback towards the central level to the detriment of feedback and management at sectoral and decentralized levels themselves (see also Valadez and Bamberger 1994).

3.2 Methodology

Methodological issues in general and ‘targets and indicators’ in particular receive considerable attention in PRSP, APR and, not surprisingly, also in independent reviews. While there are differences in quality and coverage, in our own case study of eleven countries, elaborate lists of indicators and targets were produced everywhere. Even more encouragingly, in countries that already produced more than one APR, lists get more refined from progress report to progress report. PRSP has obviously led to an upsurge in data collection, especially on poverty, and to a renewed interest in household surveys. The growing attention to qualitative, difficult-to-measure issues has also given a new impetus to methodological approaches seeking to combine qualitative and quantitative methodologies, while participatory techniques are increasingly propagated to make the conclusions more informed by the poor.

Booth and Lucas (2002) make the important point that, in setting indicators and targets, there is a strong tendency to emphasize the two extremes of the logic chain, inputs and final poverty outcomes, leaving a ‘missing middle’ in between. This constitutes a serious criticism with far-reaching implications for the quality of programmes and

M&E. The fact that different levels of indicators mentioned in PRSP or sectoral programmes are often not integrated into one causal chain similarly illustrates the absence of an underlying programme theory. Programme theory or logic models, including a process and impact theory, are powerful tools in conceptualizing interventions, monitoring and evaluation. Not surprisingly, if no prior theory can be referred to that makes clear which actions were expected to produce which results and why, it becomes extremely difficult at the time of reporting to make much sense of the findings generated (see also GTZ-BMZ 2004).

3.3 Participation of non-state actors

The overall impression from independent reviews (see, e.g., World Bank 2003b; McGee and Norton 2000; Robb 1999) is that in many countries civil society actors (CSOs) have been involved in participatory poverty assessments at the moment of preparing PRSP, but much less during subsequent phases of implementation, monitoring and evaluation. Generally, participation of CSOs is conceived of as instrumental, as a way of gathering data and insights, but within a process that is strongly government controlled. There is not much discussion in PRSP about the role that CSOs could play as an independent source of supply of M&E, nor about their capacity as actors on the M&E demand side, and this despite the participation rhetoric of the new aid paradigm. This holds even more for other actors outside the executive branch of government, such as parliament, the auditor general's office, universities, independent research institutes or national evaluation societies. While the importance of an independent general auditor's office is often stressed with respect to financial auditing, the role this office could play in performance auditing goes largely unnoticed. A similar observation can be made with respect to universities and research institutes who can add value in terms of independent and high quality data collection and particularly analysis.¹³ Some tepid progress is being made, but, ominously, the genuinely independent non-state actors that do get invited to play a role in M&E often pay for it by having their autonomy curtailed.

3.4 Quality and capacity

A good way to test for quality is to assess the annual progress and performance reports shared with donors, arguably the most tangible outputs of the national M&E systems. The superficiality of APR, particularly in terms of analytical depth, is one of the issues donors seem to agree on. APR do manage to fill in some of the gaps in baseline data that existed at the time the PRSP was first produced, and, in some cases, register changes in actual performance with baseline data, but only in exceptional cases is there an analysis of the reasons for non-achievement. While donors seem to have been taken by surprise by the sloppy quality of APR, even a superficial prior diagnosis should have alerted them to the problems that presaged such poor reporting: absence of programme theory, strong focus on monitoring to the detriment of evaluation, vaguely defined overall M&E design and unclear coordination among different actors of the system. Unpromisingly, current assessments of M&E capacity constraints and remedial action plans, both by recipient countries and donors, overwhelmingly focus on human and

¹³ Bamberger (1991) similarly notes that, given the capacities national research institutes and universities often have, it is surprising that they are so little involved in national M&E activities.

financial resources and technical issues, to the detriment of the broader institutional and policy-related issues.

4 The donor reform agenda that isn't

The hugely expanded role of recipient M&E systems under the new aid approach, even if insufficiently acknowledged in PRSP documents and the literature, is plain to anyone taking a closer look at the literature on new aid modalities. The future role for donors in M&E is, however, less clearly articulated, and some sensitive issues are not being addressed as frankly as their importance would warrant. Notwithstanding the fact that donors are setting the agenda for whatever M&E reforms are taking place in the public sector, PRSPs and APRs are fairly silent on the precise role donors are supposed to play. Similarly, even if donors probably spend more resources and energy in parallel M&E exercises than the government is able or willing to spend on national M&E, little reference is made to how such donor M&E outputs feed into the national effort. There is a major issue of ownership, or rather, lack of ownership, surrounding M&E. Donors see reform in this area clearly as an important corollary to their reduced micro-management of aid resources, and they do not hesitate to impose process conditionalities to this effect, even if sometimes in an haphazard fashion. At the same time, some are obviously uncomfortable with abandoning their own parallel M&E activities. Recipient governments on the other hand probably mostly lack a serious commitment to reform their M&E systems. They do what is expected of them, reluctantly, but they are seldom proactive, and do not correct for donor biases. An indirect proof of donor's setting of the reform agenda, and doing so badly, is the marked parallelism between the fragmentary approach they themselves adopt and similarly unbalanced development of recipient systems. In fact, donors' emphases on monitoring inputs (PEM), recently combined with a focus on monitoring final poverty outcomes and MDGs, as well as their general preference for data collection and quality, are all echoed in the particular pace and scope of recipient M&E systems reform.¹⁴

Findings summarized in the OECD-DAC harmonization and alignment survey (2005b) suggest that not much has happened on the ground. In fact, by the end of 2004, on average 30 per cent of the donor project portfolio was managed by national procedures. Desaggregation over different phases of the project cycle reveals some differences: donors seem to trust national systems slightly more when it comes to procurement (34 per cent) and are more distrustful of recipient's M&E and financial audit (both 28 per cent). Larger differences were noted among individual donors and in donor assessments of different partner countries. Similar differences among donors may be observed when it comes to engagement in new aid instruments. Donor enthusiasm for general or sector budget support and related handing over of responsibility for M&E to recipients remain very uneven: some, like Japan, France and the US have remained sceptical while others, the like-minded 'progressive' donors, including UK, the Nordics and the Netherlands, are wary of parallel donor M&E and trust as much as possible the recipient's M&E. The World Bank and the EC are also major supporters of the new aid instruments.

¹⁴ Interestingly, in its assessment of the actual state of M&E recipients' systems on the ground, the 2005 CDF evaluation contrasts improvements in sub-components as statistical data capacity with those in developing 'coordinated' monitoring systems (World Bank 2005).

For their accountability needs towards their own constituencies, budget support donors increasingly rely on the outputs, APR or sector reports now routinely produced by recipient M&E systems. This is partly a matter of principle. It may indeed be an act of good donorship on their part to give national systems time to develop, and to base their assessment more on the progress being made than on the objective quality of present monitoring and evaluation outputs. But on the other hand, they do not have much choice in this respect. These donors just cannot afford to be too openly critical of the overall quality of the national M&E systems that constitute a growing source for their own reporting without by the same token undermining the credibility of the latter. When selecting indicators for reporting, they are in a fairly comfortable position as they can stick to policy matrices they themselves had a considerable hand in negotiating. The budget support donors are, for obvious reasons, the most influential in setting the process conditionalities for national M&E reform. They tend to put most emphasis on the quality of financial reporting and on the generation of statistical data. However, after several annual rounds of APRs, donors are forced to face the fact that the analytical quality is often unacceptably low. The 2005 SPA Budget Support Alignment Survey (p. 51) has indicated, for instance, that about 71 per cent of the budget support donors consider APR unsatisfactory for their own reporting (accountability) needs.

Obviously, donors and recipient reform agendas are not independent of each other and the donors, as drivers of the reforms, are caught in a chicken-egg dilemma: as long as a minimum institutional capacity in terms of design, implementation and evaluation apparatus is not installed and functioning in a partner country, the move towards new aid instruments which shifts more responsibilities to the recipient, tends to be resisted by the more traditional donors, whose reluctance to embrace the new aid approach may be due to the fact that they face tougher than average accountability questions at home, or simply because of bureaucratic or political inertia. Whatever the reason may be, the reluctance to abandon parallel donor M&E systems is a part of the problem in that it draws attention and resources away from the institutional strengthening of recipient systems. Up to this moment in time, nobody has come up with a satisfactory solution that takes into account the differentiated donor attitudes towards the new approach to aid.

One of the issues that is not fully addressed in this respect is the question whether harmonization and alignment should be pushed as far in the field of M&E as in other fields. Handing over responsibility for the use of aid resources to the recipient, provided that a number of prior conditions are met and provided that certain process conditionalities are respected, is an approach that is difficult to fault in theory. Handing over the responsibility for providing the data and analysis that justify the use of the same resources to the same recipient is much less evident, and seems to violate the principle of independence. Recipients will always have a considerable interest to be less than frank in their assessments, to hide data that would reveal misuse of funds or lack of progress in achieving the results that were agreed upon, etcetera. Recipients thus have a strong vested interest in not setting up a transparent and competent, independent M&E system. The Paris Declaration is fairly naïve in this respect, when it repeatedly says 'until donors can rely'. This, in fact, suggests that donor-controlled M&E may be just a temporary inconvenience, an infringement of the harmonization and alignment principles that should be corrected as soon as possible. If one agrees that donors and recipient governments do not have identical agendas and interests, the donors have no other option but to insist on strong and independent monitoring and evaluation of their aid resources. National M&E systems are intrinsically unsuited for this task.

5 Which way forward?

5.1 Flying (too) high

The new approach to aid is very ambitious and it will take considerable time before it yields substantial results. This creates genuine problems for donors whose time horizons—or more precisely, those of their constituents—are much shorter. Confronted with the tension between the harmonization and alignment agenda they have endorsed, and the need to account to their own parliament, public opinion, or executive boards, donor reactions may lean in one or the other direction, but few find a middle ground. One attitude is for the donor to stick to its own parallel M&E system. The other is for the donor to become excessively lenient in assessing recipients and to see something that is in fact not there: a decent recipient M&E system in the making, just about to perform the functions of feedback and accountability in an adequate fashion and starting to live up to minimum international standards so as to also fulfil the accountability needs of donors. Donors leaning on the side of harmonization and alignment and wanting to ‘give recipient M&E a fair try’ may decide to take the interpretation of their own accountability just one step further. The new aid paradigm, with its preference for the pooling of donor resources and their blending with the recipients’ own fiscal resources, to be spent using recipient priorities and implementation systems, away from donor earmarking and micromanaging, already forced them to water down the definition of individual donor accountability. In fact, programme-based approaches make it increasingly difficult, if not impossible, to ‘assign’ final development impact to individual donor contributions, and the logical alternative is to replace it by joint donor accountability for results. Going just one step further, the donor may decide that he himself should not be assessed on final results—which are mainly the responsibility of the recipient anyway—but rather on the basis of his ‘organizational effectiveness’ in applying the principles of good donorship. While there is not yet a standardized definition for ‘organizational effectiveness’, some agencies have made efforts to render the concept more concrete (see e.g., Lehtinen 2002). Suggestions include looking at internal policies, strategies, processes, relations to partners, resource use and adaptive capability. So far, there has not been much independent research on this and not many efforts have been invested in identifying criteria or targets, except for harmonization and alignment. The Paris Declaration includes ‘harmonization and alignment’ as criteria for a donor’s aid effectiveness. This suggests that a ‘donor’s accountability towards its own constituencies’ could be defined in terms of a donor’s realization of ‘harmonization and alignment’.

The same logic is used to exercise pressure on other donors who take the other path and stick to their own M&E, donors who presumably also insist on a higher degree of earmarking and who insist on being more closely involved in implementation. Some studies (see, e.g., de Renzio et al. 2005) propose incentives to entice such reluctant donor agencies to implement the harmonization and alignment agenda, including peer control and stigmatization. The DAC Working Party on Aid Effectiveness has been urged to provide guidance on definitions and define criteria. Looking at earlier harmonization and alignment surveys, one may think about easy-to-grasp quantitative indicators including ‘number of donors missions done jointly’, ‘number of diagnostic reviews done jointly’, While one may applaud tendencies to make vague concepts operational, and monitor and assess a donor’s performance, the way in which this is done is of critical importance. The point we wish to make is that there is, in our opinion, insufficient evidence to believe that more harmonization and alignment will produce the

sought after beneficial effects in terms of poverty reduction. Until there is better evidence about the underlying explicit and implicit assumptions of the new paradigm, donors should probably err on the conservative side, and regard harmonization and alignment as the best available but, as yet, unproven means towards an end, rather than as an operational objective by which their effectiveness can be measured.

It is indeed important that donors address the reform agenda squarely, and that they do so in a concerted manner, but they should also accept that in the manner of M&E there can be too much alignment. There are too many glaring deficiencies at both the supply and demand side of national M&E that will not go away by putting the government in charge and by the occasional methodological and technical prodding, or by haphazard process conditionalities that fail to address the fundamental weaknesses. In the following section, we propose that donors adopt a more hands-on approach towards recipient's M&E, and get actively involved in the development of the so-far neglected components of the M&E system. This, we believe, will help realizing both functions of accountability and learning and benefit all stakeholders involved. Importantly, we argue that such an approach is not contradictory to the elaboration of programme-based approaches or the new aid paradigm, including alignment, on the contrary.

5.2 Seeking the middle ground

In what follows we propose a pragmatic process approach that seeks a middle position in the current debate among opposing donor camps. Between over-alignment on a recipient's M&E system that has not even been properly diagnosed on the one hand and a multitude of donor-driven parallel systems that ignore the need for institutional strengthening of national systems on the other hand, there is space for fruitful intermediate solutions.

The approach proposed here starts from the same preoccupation as the new aid paradigm: that the institutional capacity of low-income recipient countries, and in particular their M&E systems, have been undermined by uncoordinated and parallel donor-imposed rules and arrangements. Only by using the recipient 'systems', no matter their actual state, can they improve. However, given the present weaknesses on several fronts, including deep-rooted policy and institutional aspects, we believe that it is unrealistic and naïve to surmise that a combination of spontaneous dynamics from within and appropriate process conditionalities from outside will be enough to generate these improvements within a reasonable space of time (see also Booth 2005). We also feel, as argued before, that donors cannot escape their own responsibility of closely supervising national data generation and analysis. We propose a dual-track process approach emphasizing the effective building and strengthening of both supply and demand sides of recipient M&E systems through coordinated and holistic diagnosis, capacity-building, independent follow-up and process conditionality, coupled with complementary donor-managed M&E that, instead of duplicating national efforts, strives at bridging existing gaps in the national M&E system. Complementary M&E will obviously be driven by external actors, and should exclude those departments within donor agencies that are too closely involved from the M&E tasks which would require independent assessment. Importantly, the complementary M&E should be conceived and implemented in such a way that it simultaneously serves existing actual M&E needs as well as capacity-building of national demand- and/or supply-side actors. The latter may be realized by involving to the maximum degree possible—and

depending on the particular M&E function one wants to emphasize—demand- and supply-side actors in a learning-by-doing exercise. Additionally, there is an urgent need to focus on in-depth evaluations of the new aid paradigm itself. Particularly useful will be coordinated (joint) donor undertakings that do not necessarily result in unified responses, but that are nonetheless in line with the narrow definition of ‘harmonization’.¹⁵

5.3 Focus on the recipient’s system: diagnosis and capacity building

The starting point of the first track is a diagnosis of the actual state of a recipient’s M&E supply and demand. This diagnosis, identifying strengths and weaknesses, should preferably be done by a team consisting of independent M&E experts and representatives of all stakeholders involved, including donors, government and non-government actors. It serves multiple purposes:

- i) it guides donors in applying the Addis Ababa principle¹⁶ by assessing the degree to which a recipient’s M&E system can satisfy their own accountability needs and determine what complementary M&E if any is needed (see track 2);
- ii) it lays the basis for a coherent approach to capacity-building; and
- iii) it forms the basis for realistic targets to strengthen individual M&E demand and supply components, and for periodic independent assessments within a framework of process conditionality.

In order to avoid past mistakes with PEM diagnostic ‘bombardments’, a multiple stakeholder-coordinated approach, similar to the PEFA-initiative,¹⁷ should be favoured. It is worth emphasizing that a shared diagnosis need not necessarily trigger identical donor responses. The degree to which a recipient’s M&E will be considered satisfactory, for instance, will depend not only on the intrinsic quality of its supply and demand side, but also on the donor’s accountability needs towards its own constituencies, which in turn depend on the accountability approach adopted in different donor country or agency settings. This has implications for the degree and form of complementary M&E different donors will find necessary and appropriate. Different donors need not set identical targets, as long as targets are not contradictory.

¹⁵ With ‘harmonization’ is meant that donors increasingly coordinate and streamline their activities as to reduce the high transaction costs their own aid procedures inflict upon recipient governments. The OECD/DAC (2005) includes (i) establishing common arrangements, (ii) simplifying procedures and (iii) sharing of information. de Renzio (2005: 3) indicates that the concept of ‘harmonization’ is often used in a broader sense to encompass both harmonization in the narrow sense as well as alignment and ownership.

¹⁶ The Addis Ababa principle favours the use of government systems whenever possible. It reads as follows: ‘all donor assistance should be delivered through government systems unless there are compelling reasons to the contrary; where this is not possible, any alternative mechanisms or safeguards must be time-limited, and develop and build, rather than undermine or bypass, government systems’ (SPA 2001: 2).

¹⁷ The PEFA programme aims at strengthening recipient and donor ability to (i) diagnose the quality of country public expenditure, procurement and financial accountability systems and to (ii) develop a practical sequence of reform and capacity-building actions. See www.pefa.org for more information.

The choice of an appropriate diagnostic instrument will be no easy task. At this moment, donors do not really use diagnostic instruments that pass the test. Most still adopt a fairly narrow approach towards M&E, characterized by a focus on the supply side, in particular on technical and methodological issues, and not even the most crucial ones. Existing diagnostic tools and related capacity-building concentrate on input monitoring (PEM)—most donors would not even classify these as being an integral part of the M&E system—and statistical data quality, particularly that of outcome data related to the MDGs. This approach may well be understood from within donor’s own accountability needs. Recipient government accountability towards donors has traditionally been defined in terms of input management (see also Bamberger 1989, 2000; Picciotto 2003 and World Bank 2002). Driven by pressures of results-based management¹⁸ and the MDGs, a layer of ‘outcome’ accountability has been added on top. This focus on the two extremes of the causal chain has certainly not been inspired by evaluation theory, which almost unanimously emphasizes the importance of programme theory and identification of the full causal chain. Bringing this whole chain into focus would assist both learning and accountability functions, as it would allow accountabilities of the different actors to be identified at the appropriate level in the chain. Recipient government accountability should mostly be defined at the level of outputs and intermediate outcomes, as these are the issues that they have fully, or at least to a large extent, under control. Whereas one may expect from a recipient government more than just good input management, it is not realistic and counterproductive to keep recipient governments accountable for final outcomes (MDGs and poverty reduction), which are influenced by factors outside their control.

Donor diagnostics further tend to neglect policy and institutional M&E issues, and, by downplaying the principle of independence, in particular tend to overlook the M&E

Box

Available tools that might feed into the elaboration of a more holistic M&E diagnostic instrument

- The diagnostic tool for designing or reviewing a PRS monitoring system elaborated in Bedi et al. (2006).
- The readiness assessment elaborated by Kusek and Rist (2002) and the highly similar ECD diagnostic guide and action framework (see Mackay 1999).
- The checklist used for the diagnosis of PRSP M&E of five French-speaking SSA countries by the order of the Canadian International Development Agency (Lavergne and Alba 2003).
- The checklist used by Booth and Lucas (2002) in their diagnosis of PRSP M&E in 21 countries.
- The checklist used for the diagnosis of PRSP M&E in eight SSA countries by ODI’s (2003) PRSP Monitoring and Synthesis Project.
- The checklist used by Holvoet and Renard (2005) in their diagnosis of PRSP M&E of eleven SSA countries (see Table A1 in the Appendix).

¹⁸ See White (2002) for a criticism on results-based management.

demand side. Obviously, a more holistic approach towards M&E necessitates a holistic diagnostic instrument. Whereas, to the best of our knowledge, no tailor-made instruments are currently systematically used, there are some interesting donor-led and independent assessments and studies that might provide some inspiration for the elaboration of such a diagnostic instrument. The box on the preceding page gives an overview of possible inputs. Table A1 in the Appendix gives in more detail the framework that we elaborated in our desk review study.

5.4 Complementary donor-instigated M&E

Even donors that try to put the harmonization and alignment principles in practice often engage in additional M&E exercises, with diverging degrees of coordination among themselves and/or with recipient systems. Rather than regarding these as temporary aberrations on the road to true alignment, we prefer to look upon them as interesting experiments from which a lot may be learned for the second tier of the twin-track approach that we advocate. We think in particular about some performance assessment frameworks (with varying degrees of alignment to the PRSP APR and policy matrix), (joint) sector reviews, and even some isolated donor M&E of programmes. What is crucial is that such complementary M&E be well coordinated, tailored to the results of the diagnostic assessment of the recipient M&E (see track 1), and adaptive to the changing quality of the latter. Again, it is not necessary that all donors equally participate in these complementary M&E exercises; different donors, confronted with different accountability needs imposed by their own constituencies, will engage in such complementary M&E to varying degrees. What is important is that there is proper coordination among donors and with the recipient M&E system. This starts with appropriate information sharing, but it can go much beyond.

We believe that these complementary M&E exercises should go beyond delivering particular products that satisfy actual M&E needs. By involving national M&E supply- and demand-side actors where possible and appropriate, these exercises will function as learning-by-doing, a useful form of capacity-building. The involvement of national actors also increases the probability that information generated from complementary M&E feeds into the recipient's M&E. Which actors should be involved in which kind of exercises depends upon the function, accountability or learning, being emphasized. A quick look at some of the complementary M&E exercises that are currently performed on the ground suggests that donors indeed involve national actors, albeit, without always respecting basic M&E principles. Joint sector review missions (JRM) and related annual review meetings (ARM), which may be conceived as periodic assessment exercises during implementation, are currently often used as complementary M&E to annual sector progress reports. Positively, these missions are joint donor and heavily involve national stakeholders, including the implementing departments, central ministries (finance, planning) and to a lesser extent CSOs. While it is not explicitly stated in those terms, the aim of the JRM and ARM is mostly twofold, at the same time satisfying accountability and feedback needs. While there are some recent contributions (see e.g., Lehtonen 2005) to evaluation literature that indicate that there might be some ways to reconcile both functions, the tension and the need to explicitly address this issue is widely acknowledged. Standard evaluation theory and practice suggest that more often than not different functions necessitate the involvement of different actors and the use of different methodologies. In the absence of such precautions, single JRMs strive to satisfy an odd mixture of accountability and learning purposes, and typically involve

a wide range of government and non-government actors. It might be worth splitting these into two separate exercises that explicitly address one of the functions. If one wants to do a reality check on government's performance, for instance, then there is obviously a need to respect as much as possible the principle of autonomy which implies the involvement of independent actors on the demand side, such as the auditor general's office, independent policy research institutes, evaluation associations and independent CSOs. One could even argue that donors (or at least some departments in donor agencies) are themselves not independent enough to be involved in these accountability exercises.¹⁹ Additionally, accountability exercises demand careful scientific preparation. In order to be able to extrapolate findings from a limited sample, for instance, one needs to increase the external validity of the assessment as much as possible. Concretely, this implies that samples need to be representative. Failing to respect this weakens the credibility of the assessment and gives the implementing agency under scrutiny every excuse to downplay or flatly ignore 'negative' findings. On the other hand, learning necessitates zooming into and analysing unexpected results, success stories, failures or specific experiments, in the way services are delivered, for instance. Here external validity is obviously not such a major concern; on the contrary, internal validity should be optimized, so as to make credible conclusions about cause-effect linkages of the underlying programme theory. In order to identify issues under study and to ensure that feedback for future policymaking and management is optimal, one obviously needs to involve those responsible for policy and programme design and implementation. The fact that the usefulness of these 'learning' exercises potentially extends beyond the particular programme, sector, or even country setting, while the costs of such 'impact evaluation' exercises are high, justifies heavy donor involvement, at least in terms of financial and technical inputs. In order to ensure that findings from such exercises are disseminated and used internationally, one might also think about joint donor responsibility in the elaboration of international databases and in doing prospective evaluation and meta-analysis summarizing what has been learned about a particular problem or policy, what has not been learned and needs further research (see also Center for Global Development 2006).

Finally, there is a need for independent evaluation of the new aid paradigm itself, preferably led by independent (joint donor) evaluation departments, policy research institutes or universities, so as to ensure the necessary independence and to lend credibility to the findings.²⁰ To increase feedback to national policies, it makes good sense to also involve representatives of different stakeholders. In order to be useful for future policymaking, such evaluation should distinguish between process and impact evaluation. The former aims to assess whether all necessary ingredients that have been identified as necessary in the new aid paradigm have been implemented as foreseen and whether they are functioning as foreseen. Only when one is satisfied that the necessary ingredients of the new aid paradigm have been implemented, is it possible to test the new aid paradigm itself and its underlying programme theory, by checking whether the predicted final outcomes on poverty reduction are realized. Skipping the first essential phase of process evaluation mortgages any conclusion about the new aid paradigm.

¹⁹ In one revealing interview, an auditor general confided that in his view donors did not really support him in his goal to undertake independent performance auditing, because they feared that what might come out would not please them.

²⁰ One of the most ambitious efforts aiming at assessing the new aid paradigm has been the recent Joint Evaluation of General Budget Support. See IDD (2006) for an overview of results.

6 Conclusion

The present shift towards programme-based approaches requires that both donors and recipients review their M&E policies. It is expected that recipients expand and improve their presently embryonic M&E arrangements; donors are expected to harmonize and to wind down their own burdensome M&E mechanism and rely instead on recipient M&E systems. This implies a series of reforms that are technically daunting and politically sensitive, and maybe it is not so surprising that progress so far has been mitigated. Changes in recipient systems are slow and concentrated on specific components of M&E systems, biased towards the more technical and methodological components, including statistical capacities, data collection on poverty and MDGs. At the same time, there is a worrisome vagueness surrounding the overall M&E policy, particularly regarding the way the two basic principles of M&E, i.e., autonomy (or impartiality) and feedback, are guiding the institutional and organizational set-up of M&E in countries that do not necessarily excel in transparency or accountability towards the own constituencies. Donors may be partly blamed for this sorry state of affairs. If anything, they have pushed reforms that conform to their own slimmed-down and biased vision of M&E, skewed towards public finance management (input level), recently combined with a vivid interest in sophisticated data collections systems to capture the other end of the causal chain, i.e., poverty and MDG follow-up. Regarding some of the more sensitive and difficult-to-manage issues they are remarkably silent.

Nor is there reason to congratulate donors for basing their own reforms on concerted sound analysis and assorted action. In fact, most of them seem unable to escape the chicken-and-egg dilemma they are in. As long as a minimum recipient institutional capacity in terms of design, implementation and evaluation apparatus is not installed and functioning, alignment from a donor perspective is far from straightforward, and most feel compelled to duplicate recipient M&E activities, with related demands on recipient systems that go a long way to undermine the whole approach. Donor reactions are, however, far from uniform. Another group of donors fervently believes in the new aid paradigm, but gets entangled in the web of its own radical discourse. They all too readily adopt a version of 'alignment' that puts almost all the responsibility for M&E on the recipient side, notwithstanding the glaring political, institutional and technical weaknesses of partner countries. In fact, we are witnessing an unhelpful polarization of the positions on how to proceed with M&E reform. Is there, then, no intermediary position, on the one hand, between a principled handing over of M&E responsibilities to partner countries without even a prior assessment of the national system and the demand and supply pressures that shape it, and on the other hand, a continuation of non-harmonized and non-aligned donor M&E practices? This paper has sought to contribute to filling this gap and has argued that there is room for fruitful intermediate solutions. It has proposed what we believe to be a pragmatic twin-track process approach, based on a combination of effective strengthening of both demand and supply sides of recipient M&E systems through coordinated and holistic diagnosis, capacity development, independent follow-up and process conditionality, coupled with complementary externally-steered M&E that rather than duplicating national efforts strived at bridging existing gaps in the national M&E system. While it will undoubtedly take much time and energy, we believe that it is possible to move towards broad-based national ownership of M&E, without ever fully eliminating the need for donor involvement.

References

- Bamberger, M. (1989). 'The Monitoring and Evaluation of Public Sector Programs in Asia. Why are Development Programs Monitored but not Evaluated?'. *Evaluation Review*, 13 (3): 223-42.
- Bamberger, M. (1991). 'The Politics of Evaluation in Developing Countries'. *Evaluation and Programme Planning*, 14: 325-39.
- Bamberger, M. (2000). 'The Evaluation of International Development Programmes: A View from the Front'. *American Journal of Evaluation*, 21 (1): 95-102.
- Bedi, T., A. Coudouel, M. Cox, M. Goldstein, and N. Thornton (2006). *Beyond the Numbers. Understanding the Institutions for Poverty Reduction Strategies*. Washington, DC: World Bank.
- Booth, D. (2005). 'Missing Links in the Politics of Development: Learning from the PRSP Experiment'. ODI Working Paper No. 256. London: Overseas Development Institute.
- Booth, D., and H. Lucas (2002). 'Good Practice in the Development of PRSP Indicators and Monitoring Systems'. ODI Working Paper No. 172. London: Overseas Development Institute.
- Boyle, R., D. Lemaire, and R.C. Rist (1999). 'Introduction: Building Evaluation Capacity'. In R. Boyle and D. Lemaire (eds), *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick and London: Transaction Publishers.
- Center for Global Development (2006). *When Will we ever Learn? Improving Lives through Impact Evaluation*. Report of the Evaluation Gap Working Group. Washington, DC: Center for Global Development.
- De Renzio, P., D. Booth, A. Rogerson, and Z. Curran (2005). 'Incentives for Harmonization and Alignment in Aid Agencies. ODI Working Paper No. 248. London: Overseas Development Institute.
- Eriksson, J. (2001). *The Drive to Partnership. Aid Coordination and the World Bank*. Washington, DC: World Bank Operations Evaluation Department.
- Evans, A., and A. van Diesen (2002). 'Tanzania's Poverty Monitoring System. A Review of Early Experience and Current Challenges'. London and Dar es Salaam.
- Evans, A., and E. Ngalwea (2001). 'Institutionalizing the PRSP Approach in Tanzania'. In D. Booth (ed.), *The PRSP Institutionalization Study: Final Report*. London: Overseas Development Institute.
- GTZ-BMZ (Deutsche Gesellschaft für Technische Zusammenarbeit–German Federal Ministry for Economic Cooperation and Development) (2004). *National Monitoring of Strategies for Sustainable Poverty Reduction/PRSPs*, vol. 1. Eshborn: GTZ, Governance and Democracy Division, Mainstreaming Poverty Reduction Project.
- Hauge, A. (2001). 'Strengthening Capacity for Monitoring and Evaluation in Uganda: A Results-based Management Perspective'. ECD Working Paper Series No. 8. Washington, DC: World Bank Operations Evaluation Department.

- Hauge, A. (2003). 'The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda'. ECD Working Paper Series No. 10. Washington, DC: World Bank Operations Evaluation Department.
- Holvoet, N., and R. Renard (2005). 'Putting the New Aid Paradigm to Work: Challenges for Monitoring and Evaluation'. IDPM-UA Discussion Paper No. 2. Antwerp: Institute of Development Policy and Management, University of Antwerp.
- Holvoet, N., and R. Renard (2007). 'Monitoring and Evaluation under the PRSP: Solid Rock or Quicksand?'. *Evaluation and Programme Planning*, 30: 66-81.
- IDD (International Development Department and Associates) (2006). *Evaluation of General Budget Support: Synthesis Report*. Birmingham: IDD and Associates.
- IMF (2004). *Report on the Evaluation of Poverty Reduction Strategy Papers (PRSPs) and the Poverty Reduction and Growth Facility (PRGF)*. Washington, DC: IMF Independent Evaluation Office.
- IMF and IDA (2003). *Poverty Reduction Strategy Papers – Detailed Analysis of Progress in Implementation*. Washington, DC: IMF and World Bank (IDA).
- IMF and World Bank (2005). *2005 Review of the Poverty Reduction Strategy Approach: Balancing Accountabilities and Scaling Up Results*. Washington, DC: IMF and World Bank.
- Kusek, J. Z., and R. C. Rist (2002). 'Building Results-Based Monitoring and Evaluation Systems: Assessing Developing Countries Readiness'. *Zeitschrift für Evaluation*, 1: 151-58.
- Kusek, J. Z., and R. C. Rist (2004). *Ten Steps to a Results-Based Monitoring and Evaluation System. A Handbook for Development Practitioners*. Washington, DC: World Bank.
- Lavergne, R., and A. Alba (2003). *CIDA Primer on Programme-Based Approaches*. Gatineau: CIDA (Canadian International Development Agency)—Analysis and Research Division, Policy Branch.
- Lehtinen, T. (2002). 'Measuring the Performance of EC Development Cooperation: Lessons from the Experiences of International Development Agencies'. ECDPM Discussion Paper No. 41. Maastricht: European Centre for Development Policy Management.
- Lehtonen, M. (2005). 'OECD Environmental Performance Review. Accountability (f)or Learning', *Evaluation*, 11 (2): 169-88.
- Lucas, H., D. Evans, and K. Pasteur with R. Lloyd (2004). 'Research on the Current State of PRS Monitoring Systems'. IDS Discussion Paper No. 382. Sussex: Institute of Development Studies.
- Mackay, K. (1999). 'Evaluation Capacity Development: A Diagnostic Guide and Action Framework'. ECD Working Paper Series No. 6. Washington, DC: World Bank.
- May, E., D. Shand, K. Mackay, F. Rojas, and J. Saavedra (eds) (2006). *Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank/Inter-American Development Bank Conference*. Washington, DC: World Bank.

- McGee, R., and A. Norton (2000). 'Participation in Poverty Reduction Strategies: a Synthesis of Experience with Participatory Approaches to Policy Design, Implementation and Monitoring'. IDS Working Paper No. 109. Sussex: Institute of Development Studies.
- ODI (Overseas Development Institute) (2003). 'PRS Monitoring in Africa'. PRSP Synthesis Note No. 7. London: PRSP Monitoring and Synthesis Project-ODI.
- ODI (Overseas Development Institute) (2004). 'PRSP Annual Progress Reports and Joint Staff Assessments: A Review of Progress'. PRSP Briefing Note No. 9. London: PRSP Monitoring and Synthesis Project-ODI.
- OECD-DAC (2003). *Harmonizing Donor Practices for Effective Aid Delivery*, vol.1. Paris: OECD- Development Assistance Committee.
- OECD-DAC (2005a). *Harmonizing Donor Practices for Effective Aid Delivery*, vol. 2. Paris: OECD-Development Assistance Committee.
- OECD-DAC (2005b). *Survey on Harmonization and Alignment*. Paris: OECD-Development Assistance Committee.
- OECD-World Bank (2005). *Harmonization, Alignment, Results: Progress Report on Aid Effectiveness*. Report of the High Level Forum, 28 February-2 March. Paris: OECD Publications.
- Picciotto, R. (2003). 'International Trends and Development Evaluation: The Need for Ideas', *American Journal of Evaluation*, 24 (2): 227-34.
- Prennushi, G., G. Rubio, and K. Subbarao (2001). 'Monitoring and Evaluation plus Annexes'. In *PRSP Sourcebook*. Washington, DC: World Bank, 107-30 and 433-62.
- Rist, R. C. (ed.) (1990). *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations*. New Brunswick, NJ: Transaction Publishers.
- Rossi, P. H., H. E. Freeman, and M. W. Lipsey (2004). *Evaluation: A Systematic Approach*, 7th edition. Thousand Oaks: Sage.
- SPA (Special Partnership with Africa) (2001). 'Conclusions of the Joint Meeting of the ECA PRSP Learning Group and the SPA Technical Group'. 7 November. Addis Ababa. Available at: www.poverty2.forumone.com/files/9725_spa4.pdf .
- SPA-6 Budget Support Working Group (2005). *Survey of the Alignment of Budget Support and Balance of Payments Support with National PRS Processes*. Report by the BSWG Co-Chairs). Available at: www.spa-psa.org/index.jsp?sid=1&id=928&pid=1210
- Valadez J., and M. Bamberger (1994). 'Organizational and Management Issues in Programme Evaluation'. In *World Bank, Monitoring and Evaluating Social Programmes in Developing Countries: A Handbook for Policymakers, Managers and Researchers*. Washington, DC: World Bank, 403-42.
- White, H. (2002). 'The Road to Nowhere? Results-based Management in International Cooperation'. *Evaluating International Cooperation Think Tank*, No. 3. Available at: www.oneworld.org/thinktank/evaluate/edit3.htm .

- World Bank (2002). '2002 Annual Report on Evaluation Capacity Development: 2002 Annual Report'. Washington, DC: World Bank Operations Evaluation Department.
- World Bank (2003a). *Toward Country-led Development. A Multi-Partner Evaluation of the Comprehensive Development Framework. Synthesis Report.* Washington, DC: World Bank.
- World Bank (2003b). *Participation in Monitoring and Evaluation of PRSPs. A Document Review of Trends and Approaches Emerging from 21 Full PRSPs.* Washington, DC: World Bank Social Development Department.
- World Bank (2004). *The Poverty Reduction Strategy Initiative. An Independent Evaluation of the World Bank's Support Through 2003.* Washington, DC: World Bank Operations Evaluation Department.
- World Bank (2005). *Enabling Capacity to Achieve Results – 2005 Comprehensive Development Framework (CDF) Progress Report.* Washington. DC: World Bank Operations Policy and Country Services.

Appendix

Table A1
The checklist used by Holvoet and Renard (2005, 2007)
in their diagnosis of PRSP M&E in eleven SSA countries

Topics	Questions
I. Policy	
1 The evaluation plan	Is there a comprehensive evaluation plan, indicating what to evaluate, why, how, for whom?
2 M versus E	Are the difference and the relationship between M and E clearly spelled out?
3 Autonomy & impartiality (accountability)	Is the need for autonomy and impartiality explicitly mentioned? Does the M&E plan allow for tough issues to be analysed? Is there an independent budget?
4 Feedback	Is there an explicit and consistent approach to reporting, dissemination, and integration?
5 Alignment planning & budgeting	Is there integration of M&E results in planning and budgeting?
II. Methodology	
6 Selection of indicators	Is it clear what to monitor and evaluate? Is there a list of indicators?
7 Selection criteria	Are the criteria for the selection of indicators clear? And who selects?
8 Priority setting	Is the need acknowledged to set priorities and limit the number of indicators to be monitored?
9 Causality chain	Are different levels of indicators (input-output-outcome-impact) explicitly linked (programme theory)? (vertical logic)
10 Methodologies used	Is it clear how to monitor and evaluate? Are methodologies well identified and mutually integrated?
11 Data collection	Are sources of data collection clearly identified? Are indicators linked to sources of data collection? (horizontal logic)
III. Organization	
12 Coordination & oversight	Is there an appropriate institutional structure for coordination, support, central oversight, and feedback? With different stakeholders?
13 Statistical office	Are surveys, censuses etc streamlined into M&E needs? Is the role of the statistical office in M&E clear?
14 Line ministries	Are there M&E units in line ministries and semi-governmental institutions (parastatals), and are these properly relayed to central unit?
15 Decentralized levels	Are there M&E units at decentralized levels and are these properly relayed to central unit?
16 Link with projects	Is there any effort to relay with/coordinate with donor M&E mechanisms for projects?
IV. Capacity	
17 Problem acknowledged	Are current weaknesses in the system identified?
18 Capacity-building plan	Are there plans for remediation? Do these include training, appropriate salaries, etc.

Table A1 continues

Table A1 (con't)
 The checklist used by Holvoet and Renard (2005, 2007)
 in their diagnosis of PRSP M&E in eleven SSA countries

Topics	Question
V. Participation of actors outside government	
19 Parliament	Is the role of Parliament properly recognized, and is there alignment with Parliamentary control and oversight procedures?
20 Civil society	Is the role of civil society recognized? Are there clear procedures for the participation of civil society? Is the participation institutionally arranged or rather ad-hoc?
21 Donors	Is the role of donors recognized? Are there clear procedures for participation of donors?
VI. Quality	
22 Effective use of M&E in APR	Is there a presentation of relevant M&E results? Are results compared to targets? Is there an analysis of discrepancies?
23 Internal usage of APR	Is the APR also used for internal purposes? Is it an instrument of national policy-making and/or policy-influencing and advocacy?