

Taking Stock of Monitoring and Evaluation Arrangements in the Context of Poverty Reduction Strategy Papers: Evidence from 20 Aid-Dependent Countries in Sub-Saharan Africa

Nathalie Holvoet, Marie Gildemyn and Liesbeth Inberg*

Shifts in thinking and practice concerning aid have posed serious challenges in monitoring and evaluation (M&E) for both donors and recipients, but progress in reform remains slow. An important first step in any initiative involving M&E capacity development is the diagnosis of the systems' current status. This article presents a diagnostic checklist that captures issues of M&E policy: indicators, data collection and methodology; organisation; capacity-building; participation of non-governmental actors; and use. It applies it to a review of the PRSP M&E arrangements of 20 aid-dependent countries in sub-Saharan Africa to demonstrate comparative strengths and weaknesses.

Key words: Monitoring and evaluation systems, reform, Poverty Reduction Strategy Papers (PRSP), Sub-Saharan Africa

1 Introduction

Since the turn of the century, shifts in aid policies and practices have been propagated with the aim of improving aid and development effectiveness on the ground. The 2005 Paris Declaration (OECD/DAC, 2005) and the 2008 Accra Agenda for Action (3rd High-Level Forum, 2008) specify the reform agenda for donors and recipients around five key principles (ownership, harmonisation, alignment, management for development results and mutual accountability), translating them into a set of twelve follow-up indicators. Recipients are asked to elaborate sound comprehensive policies (Poverty Reduction Strategy Papers or PRSPs) aimed at reducing poverty and based upon broad participation from both governmental and non-governmental actors. They must also build and reinforce results-oriented frameworks and arrangements for planning, budgeting, implementation, monitoring and evaluation. Donors are expected to share information with each other, harmonise their aid-management procedures and align them as much as possible with the policy and institutional frameworks of recipient countries. The aid modalities that most obviously correspond to this evolution in aid thinking are sector and general budget support (GBS), in which donors disburse aid

*Institute of Development Policy and Management, University of Antwerp, Lange St. Annastraat 7, 2000 Antwerp, Belgium (nathalie.holvoet@ua.ac.be).

money directly through sector or national budgets, combined with policy dialogue, technical assistance and well-aligned pilot projects aimed at improving policies and procedures that should ultimately lead to poverty reduction on the ground.

This approach has numerous implications in the area of monitoring and evaluation (M&E). The recognition that 'learning' and 'accountability' are crucial ingredients for the realisation of several of the Paris principles has renewed interest in M&E. The general reform agenda also applies to M&E and demands considerable shifts in the M&E policies and practices of both donors and recipients (Thomas, 2010). While recipients are urged to build and strengthen their M&E systems in order to satisfy the need for learning and accountability, donors are expected to scale down their own parallel M&E arrangements, or at least harmonise with each other in order to reduce the transaction costs for the M&E arrangements of recipient countries.

On the ground, progress in the reform agenda is slow and uneven. The 2006, 2008 and 2011 Monitoring Surveys of the Paris Declaration reveal that donors still rely overwhelmingly on their own institutional apparatus; they are slow to harmonise with other donors, and they hardly ever align with the M&E apparatus of recipient countries, which they consider weak (OECD/DAC, 2007: 33-4, 2008: 20; OECD, 2011: 89-90). While the 2011 survey shows considerable progress in the number of partner countries which have results-oriented frameworks that were deemed adequate (from 2 out of 29 in 2006 and 3 out of 54 in 2008 to 15 out of 76 in 2011), the 2010 target of 36% has not been reached (OECD/DAC, 2007: 35, 2008: 58-59; OECD, 2011: 86). This disappointing state of affairs on the ground is not entirely surprising: the M&E reform agenda is clearly trapped in a 'chicken-and-egg' dilemma. In times when donors are facing pressure because of increasing demands for accountability from their own constituencies and are tending to adopt a more risk-averse attitude, they are not eager to rely upon recipient M&E systems. At the same time, such reluctance to align and at least make use of existing components hampers processes of elaboration and maturation for the M&E systems of recipients.

In a similar vein, the 2010 OECD/DAC review of the ways in which aid agencies are adapting their M&E systems to the changing aid architecture describes how donors are facing pressure to involve partner countries more actively, even as issues are emerging with regard to deficiencies in the M&E independence and capacity of recipient countries. Against this background, aid agencies are increasingly acknowledging the importance of developing M&E capacity in recipient countries. To date, however, there seems to be relatively little strategic engagement in this area, even amongst those aid agencies that mention it in their mandates. The need for further guidance in this field has been particularly emphasised by the M&E units of aid agencies, which are facing a continuously expanding portfolio of activities, while struggling to come to grips with their own persistent capacity-related weaknesses (OECD/DAC, 2010).

This article aims to contribute to this challenging and policy-relevant research agenda, which has thus far remained underexplored in the academic literature. It starts from the assertion that, regardless of the approach adopted, an important first step in any M&E capacity-building effort is to take stock of what already exists. This is consistent with research on public-sector reform in low-income, aid-dependent

countries (Santiso, 2008) and the institutional-change literature (North, 1990) which has highlighted that small incremental changes to existing systems might be more feasible and workable than radical and abrupt changes that seek to impose blueprints from outside. To date, however, little systematic evidence has been available regarding the current outlook of M&E systems beyond the World Bank Comprehensive Development Framework (CDF) data, which are also used in the Paris Declaration reviews. While the CDF 'results-focus' indicator provides comparable data on a large sample of countries (World Bank, 2007), it is restrictive in the dimensions it captures, focusing specifically on three sub-components (namely, the quality of available information, the accessibility to relevant stakeholders and the existence of a coordinated country-level M&E system). In order to provide a more comprehensive picture of M&E systems, our review of 20 aid-dependent countries in sub-Saharan Africa is based upon a checklist that captures issues related to M&E policy: indicators, data collection and methodology; organisation; capacity-building; use; and participation of non-governmental actors. We draw upon a combination of quantitative and qualitative assessments to provide an overview of how core elements of the evaluation function are covered, with the aim of informing (joint) efforts to strengthen M&E systems. The next section presents details on the methodology used in this desk-based study. Findings are presented and discussed in Section 3 while Section 4 concludes.

2 Taking stock of PRSP monitoring and evaluation systems

2.1 Towards a diagnostic checklist

To the best of our knowledge, none of the currently available checklists or assessment tools allows a comprehensive overview of the quality of M&E arrangements used in the context of PRSPs. Although each country must define the grid and outlook of its own system, a number of basic M&E principles and guidelines should be followed in order to avoid loopholes in the system. Our own checklist draws upon documents that specify these principles and guidelines, including the African Evaluation Guidelines (Nairobi Network et al., 2002) and the OECD/DAC Principles for the Evaluation of Development Assistance (OECD/DAC, 1991). We also reviewed meta-evaluation literature (for example, Stufflebeam, 1974), along with several existing M&E assessment tools, including the OECD/DAC working tool for peer reviews and assessments (OECD/DAC, 2006), the diagnostic tool for the assessment of the institutional dimensions of PRS monitoring systems (Bedi et al., 2006), the Evaluation Capacity Development Diagnostic Guide and Action Framework (Mackay, 1999) and the related Kusek and Rist Readiness Assessment (Kusek and Rist, 2004). Our checklist (see Annex 1) includes 23 questions, sub-divided over 6 broad M&E dimensions : (i) policy, (ii) indicators, data collection and methodology, (iii) organisational issues, (iv) capacity-building, (v) participation of non-governmental actors and (vi) use. PRSP M&E arrangements were rated on each of the 23 items using an ordinal four-point scale (weak, partially satisfactory, satisfactory, excellent). The quantitative assessment allows the aggregation of findings over several countries, drawing a picture of comparative

strengths and weaknesses, without using it to rank countries scientifically. Each of the ratings was complemented and justified with qualitative comments.

Although we find the checklist valuable and also applicable to take stock of M&E systems at sector level,¹ we are aware of its limitations. First, while the inter-item correlation for the entire checklist is relatively strong (Cronbach's alpha 0.835), the inter-item correlation for each of the six areas is not consistently high, with scores ranging from 0.767 for 'evaluation policy' to 0.136 for 'participation of non-governmental actors'. While the analysis of findings is structured along the six dimensions, we also highlight diverging results amongst items classified under the same heading. Second, while this desk-based study is useful for analysing and comparing data from a substantial number of countries, it does not allow us to capture the subtleties of individual M&E systems. In a further stage, the aim would be to complement this study with fieldwork for a limited number of cases, using political-economy analysis to explore in more detail the underlying institutional and political processes that are at play and analyse in more depth localised informal practices.

2.2 Sample and data source

This desk-based cross-national study encompasses all 20 countries in sub-Saharan Africa (SSA) that had at least finalised a second Poverty Reduction Strategy Paper (PRSP) as of May 2010. These countries are as follows (in alphabetical order): Benin, Burkina Faso, Cameroon, Cape Verde, Ethiopia, Gambia, Ghana, Guinea, Madagascar, Malawi, Mali, Mauritania, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, Tanzania, Uganda and Zambia. The choice for PRSP countries is based upon the fact that PRSPs are considered important country-owned policy documents. They specify policy priorities, which are translated into a matrix of measurable indicators and targets. They also contain information on budgeting, implementation, monitoring and evaluation. Once approved by the donor community, the PRSP provides the framework for international donor engagement. The recipient country must prepare an Annual Progress Report (APR), which describes and analyses the progress achieved on the various targets, drawing on information from M&E. In order to respect the process and iterative character of PRSPs and to ensure the elaboration of at least one APR, we limited our sample to countries that had at least entered the second PRSP round. Finally, we focus on SSA, as most of the PRSP countries are situated in that region.

For each of the 20 countries, we reviewed the same set of official PRSP documents, including the most recent PRSP and APRs (see <http://www.worldbank.org/prsp>). We also reviewed Joint Staff Advisory Notes (JSAN), which provide comments by IMF and World Bank staff with regard to the PRSPs and APRs. The discussion below makes explicit reference to any useful information on M&E that was provided. Our sample of 20 countries also includes the 11 countries that

1. The scope of a sector diagnosis is obviously more limited but key components and guiding principles of a sector M&E system largely overlap with those of a central M&E system. An important specific issue within a sector diagnosis is the linkage of sector M&E systems to the central M&E system (Mackay, 2006). For a sector-specific format of the M&E checklist and an application to health sector M&E arrangements in Rwanda and Uganda, see Holvoet and Inberg (2011; 2012).

were included in a similar review in 2005 (Holvoet and Renard, 2007). While the focus of this study does not constitute a comparison of the two assessment exercises, the discussion below does introduce a comparative perspective for this sub-sample of countries whenever relevant.

2.3 Procedure

In order to decrease subjectivity, three researchers participated in this diagnostic review. First, two researchers independently graded the 20 M&E systems on the 23 items using the same document base (for a similar methodology, see Isham et al., 1995). To justify the grading, the two researchers complemented each quantitative assessment with qualitative information, which took approximately four to five hours per country. The third (lead) researcher compiled the quantitative and qualitative assessments, checked the inter-rater reliability and identified issues that called for further discussion. Although there was a high degree of consensus amongst the researchers (intra-class correlation coefficient = 0.72 [$p < 0.001$], 95% CI [0.664, 0.767]), there were also some points of disagreement. These points were discussed during a calibration meeting, according to which the ratings were finalised by the lead researcher and endorsed by the other two. The research findings were then analysed independently by the three researchers, drawing upon the qualitative assessments and insights from previous desk-based and field studies on the topic. Finally, analyses were cross-checked, discussed and finalised.

3 Findings and discussion

Table 1 provides an overview of the assessment results, highlighting for each of the items the number and percentage of countries that scored ‘weak’, ‘partially satisfactory’, ‘satisfactory’ and ‘excellent’.

Table 1: Review of the PRSP M&E systems of 20 aid-dependent countries in SSA

Items	Weak	Partially satisfactory	Satisfactory	Excellent
I. Policy				
1. Evaluation plan	4 (20%)	10 (50%)	4 (20%)	2 (10%)
2. M versus E	9 (45%)	6 (30%)	4 (20%)	1 (5%)
3. Autonomy & impartiality	17 (85%)	3 (15%)	0	0
4. Reporting & dissemination	3 (15%)	10 (50%)	5 (25%)	2 (10%)
5. Alignment with planning & budgeting	10 (50%)	6 (30%)	4 (20%)	0
II. Indicators, data collection and methodology				
6. Selection of indicators	1 (5%)	3 (15%)	4 (20%)	12 (60%)
7. Selection criteria	13 (65%)	2 (10%)	2 (10%)	3 (15%)
8. Priority setting	8 (40%)	8 (40%)	4 (20%)	0

9.Causality chain	9 (45%)	5 (25%)	5 (25%)	1 (5%)
10.Methodologies used	7 (35%)	10 (50%)	3 (15%)	0
11.Data collection	5 (25%)	6 (30%)	8 (40%)	1 (5%)
III. Organisation				
12.Co-ordination & oversight	0	15 (75%)	5 (25%)	0
13.Statistical office	4 (20%)	8 (40%)	8 (40%)	0
14.Line ministries	4 (20%)	11 (55%)	5 (25%)	0
15.Decentralised levels	6 (30%)	13 (65%)	1 (5%)	0
16.Link with projects	18 (90%)	2 (10%)	0	0
IV. Capacity-building				
17. Capacity diagnosis	5 (25%)	7 (35%)	6 (30%)	2 (10%)
18. Capacity-building plan	2 (10%)	14 (70%)	1 (5%)	3 (15%)
V. Participation of non-governmental actors				
19.Parliament	13 (65%)	4 (20%)	3 (15%)	0
20.Civil society	3 (15%)	12 (60%)	5 (25%)	0
21.Donors	7 (35%)	10 (50%)	3 (15%)	0
VI. Use				
22.Use in Annual Progress Reports	2 (10%)	15 (75%)	3 (15%)	0
23. Use within country	11 (55%)	9 (45%)	0	0

Source: Authors' findings and analysis

In general, our findings support evidence from other related research (Holvoet and Renard, 2007; Holvoet and Rombouts, 2008; Wood et al., 2008; World Bank, 2005, 2007), in that M&E systems generally score better on technical components (and more particularly on indicators, data collection and statistics development) and somewhat weaker on policy and organisational issues, which are more difficult to address. The discussion below is structured along the six broad areas of M&E; it suggests the most important patterns in our findings, supplementing the quantitative assessment with insights from qualitative analysis.

3.1 Policy

While the majority of the countries in the sample are undertaking efforts to elaborate **evaluation plans**, only six countries provided comprehensive overviews of what they are evaluating, why, how and for whom. Interestingly, the quality of the evaluation plan appears to be a good proxy for the overall quality of M&E activities. Ghana and Tanzania, which score high on most of the 23 items, also have the most developed evaluation plans; the other four countries whose evaluation plans are considered satisfactory (Uganda, Zambia, Rwanda and Mozambique) also perform relatively well on most of the other dimensions. The reverse is also true: countries whose evaluation plans are relatively weak (for example, Cape Verde, Guinea, Madagascar) tend to have lower scores. Our review also confirms the previously noted bias towards 'statistics' in M&E development; most countries that have not (or not yet) developed overall comprehensive evaluation plans (for example, Cameroon, Cape Verde, Guinea, Mauretania, Niger, Senegal, Sierra Leone) do have strategies for developing statistics.

This is consistent with evidence from the 2005 CDF evaluation and the 2008 Evaluation of the Paris Declaration, which highlight substantial improvements in statistical data capacity in cases where the rest of the system tends to be in the embryonic stages (Wood et al., 2008; World Bank, 2005).

About half of the countries do not make any **distinction between monitoring and evaluation** in the PRSP and use the notions interchangeably. Several countries (for example, Ghana, Mauritania, Niger, Rwanda) clearly specify the difference between ‘monitoring’ and ‘evaluation’, although they fail to explain how ‘monitoring’ will feed into ‘evaluation’ or how linkages will be established amongst the various entities responsible for monitoring and evaluation. If anything, the emphasis remains predominantly on monitoring, as was also the case in our 2005 desk-based study. While a focus on monitoring may be a logical first step within the context of a sequencing approach² towards the setting-up of an M&E system, there is nowadays also a general tendency for monitoring to crowd-out evaluation. This phenomenon has, amongst others, been observed within aid agencies where evaluation has come under stress as a result of an increased focus on performance monitoring (see Liverani and Lundgren, 2007).

Particularly when the focus is limited to ‘aggregate’ indicators, monitoring might also be less politically sensitive than evaluation. The more analytical evaluative exercises tend to make underlying causal chains explicit and to allow distinctions between deficient implementation and conceptual flaws in the design of policies, or between ‘doing things right’ and ‘doing the right things’ (Kusek and Rist, 2004). Although a government that is committed to inclusive development and poverty reduction should ideally be motivated to learn in order to improve its policies and outcomes, it would be naïve to assume that political leaders and bureaucrats are always eager to learn from evidence. Investment in evaluative exercises and ‘knowledge’ can even become counterproductive when substantial amounts of budget support can be generated without strong ‘poverty-reducing credentials’. As Pritchett (2002: 268) observes, ‘if a program can already generate sufficient support to be adequately funded then knowledge is a danger’.

In a similar vein, M&E chapters in PRSPs remain remarkably silent on **autonomy and impartiality**, two key M&E principles that are vital to the realisation of ‘accountability’, which is one of the two M&E core objectives. The only countries that to some extent recognise the importance of autonomy and impartiality are Guinea, Tanzania and Zambia. None of the countries, however, discusses the implications of ‘independence’ in terms of the scope, set-up and location of monitoring and evaluation units, appointments and financial and human resources. This is not completely surprising: ‘independence’ and ‘impartiality’ are particularly challenging in many of the SSA countries under study, where democratic checks and balances are often curtailed (see, for example, Van de Walle, 2005). Authoritarian regimes are particularly unlikely to stimulate the accountability aspect of M&E, as doing so could directly jeopardise their power position. As indicated by Gordillo and Andersson (2004: 309), ‘the more dramatic the power asymmetries amongst actors – that is, the more power the political

2. The importance of ‘sequencing’ has particularly been discussed in the context of the development of Public Finance Management (PFM) systems (see Bietenhader and Bergmann, 2010).

elites possess relative to the power of citizens and their political opposition and other actors – the weaker the incentive is for such an elite to transform established M&E programmes into an instrument for more public sector accountability’. Actions undertaken to curtail critical analysis and insight range from overt forms of repression to more covert forms, including impediments to the autonomous functioning of monitoring and (particularly) evaluation units through restrictions in terms of location and financial and human resources. This setting, however, is not the only one in which ‘independence’ is a sensitive issue. A 2010 OECD/DAC review on evaluation in donor agencies cites improvements in strengthening the independence of the evaluation units, although it simultaneously stresses that these improvements are only of recent origin, have been associated with major conflicts and do not apply in the areas of selection of evaluation objectives (OECD/DAC, 2010: 25).

Finally, the possibility of tension between ‘accountability’ and the other M&E core objective of ‘learning and feedback’ is not addressed in any of the documents. The acknowledgement of a possible trade-off is particularly important when deciding upon the institutional location of monitoring and evaluation unit(s). As Valadez and Bamberger (1994) highlight, there is no ideal set-up and different scenarios are possible (including one M&E unit or separating monitoring and evaluation units; locating closer or further away from implementation), but each set-up has different implications in terms of satisfying goals of ‘accountability’ and ‘learning and feedback’. Whereas accountability necessitates independence and autonomy, ‘learning and feedback’ is bolstered by a certain degree of functional integration into decision-making, policy and operational arenas, or at least by early buy-in of different stakeholders.

With regard to feedback, most of the countries in the sample have elaborated plans or strategies for the **reporting and dissemination** of M&E findings. Several countries (for example, Ghana, Mozambique, Sierra Leone, Tanzania, Zambia) have adopted a variety of dissemination strategies (for example, using the media and simplified M&E reports in various local languages) in order to target different audiences. Results are more modest with regard to the **integration of M&E results into planning and budgeting**. If anything, the discourse on results-based budgeting and management and evidence-based³ planning and policy-making has clearly influenced the documents. Although the large majority of the sample countries refer to these ‘good practices’, they simultaneously highlight that these practices are not (or not yet) functional and are difficult to implement. This is not surprising, given the fact that the move towards ‘evidence-based’ planning and policies is difficult everywhere, as is the confrontation of ‘real’ with ‘financial’ spheres in the context of performance-oriented budgeting (see Pollitt, 2001). One of the obvious obstacles is the serious time lag in the production of

3 Although the idea that evidence should inform policy-planning and -making is widely accepted, the debate on what exactly constitutes credible evidence is far from being settled (Donaldson et al., 2008). In the context of this article, the following definition of evidence is adopted: ‘output from more formal and systematic enquiries generated by government departments, research institutions, universities, charitable foundations, consultancy organizations, and a variety of agencies and intermediaries’ (Davies et al., 1999: 3). In none of the documents reviewed was the term ‘evidence-based’ defined; however, the type of evidence most often referred to was statistical data, output of surveys and management information systems, and in some cases evidence gathered through participatory and qualitative research.

the APRs, even in such countries as Uganda, which is generally doing better. As stated in the JSAN, 'the APR took more than a year to prepare, and by the time it was completed it was somewhat out-of-date' (IDA and IMF, 2008: 10). Ghana and Rwanda are exceptional cases in this respect. In Ghana, recommendations from the APRs have influenced the annual national budgets and the formulation of the second PRSP (Republic of Ghana, 2005: 70). In Rwanda, there is a move towards the alignment of performance results with planning and budgeting in some sectors (Republic of Rwanda, 2007). Comparison of results from the 2005 and 2010 reviews reveals no major improvements over time. In contrast, in Mozambique, for instance, the division of the former Ministry of Finance and Planning into two separate ministries seriously impairs the M&E feedback function.

3.2 Indicators, data collection and methodology

More than half of the countries have a score of 'excellent' on the **selection of indicators**, indicating that this is by far the most developed aspect of the M&E arrangements in the 20 countries under review. Only one country has a score of 'weak' and another 'partially satisfactory'; and there is also evidence of progress over time in this dimension. The JSAN of Senegal's second PRSP, for instance, explicitly highlights the progress in the identification of monitoring indicators from the first to the second PRSP (IMF and IDA, 2006).

More than half of the countries are not transparent on the underlying **selection criteria**, however, and they are even less transparent with regard to the stakeholders involved in the selection process. Nine of the 20 countries have very long lists of indicators (500 in the case of Ethiopia), which suggests the need for further **prioritisation** and streamlining. There are also good practices, as in the case of Burkina Faso, which distinguishes amongst a limited set of core macro-policy indicators (47 indicators) and additional sector and other indicators (Ministry of Economy and Development of Burkina Faso, 2004). Not all countries are necessarily improving on this dimension over time. While Tanzania prioritised amongst the indicators in its first M&E plan, the second plan includes even more indicators and contains no explicit acknowledgement of the need for prioritisation.

On a critical note, prioritisation is obviously not without danger, particularly if it implies a unique focus on 'aggregates', as this may easily obscure distortions at disaggregated levels (Ravallion, 2006). This is particularly relevant in the context of general budget support, in which donors increasingly move away from implementation realities on the ground and where improvement in aggregates may easily conceal inequalities along layers of region, sex, age or ethnicity. Evaluative exercises such as Poverty and Social Impact Analysis (PSIA), which aim to examine the likely impact of reforms on poverty and social development within various groups in society (especially the poor and most vulnerable; see World Bank, 2003), can be particularly valuable in this respect.

One of the most important issues with vast implications for the feasibility and quality of evaluation is the degree to which various indicators have been integrated into a **logic causal chain** of inputs, outputs, outcomes and impact. The articulation of the causality chain implies the presence of some sort of underlying programme theory

during the formulation of programmes and theories, which subsequently improves their 'evaluability'. There seems to be a marked difference in this regard between countries that have longer experience with the PRSP process and newcomers, as the majority of the 11 countries included in the first review have improved on this dimension. Although this was one of the weaker dimensions in the 2005 review, with none of the countries scoring 'satisfactory', Ghana now has a score of 'excellent', specifying the full causality chain for each of the 138 core indicators, while four other countries are doing so for at least some of the indicators. This is in sharp contrast to countries that have less PRSP experience, in which indicators are at best differentiated over different levels and are hardly ever integrated into a causal chain. Rwanda, which is the best overall performer in the area of indicators, data collection and methodology, forms an exception to this rule, having provided a detailed description of the causal chain in explaining results, accompanied by illustration through concrete examples (Republic of Rwanda, 2007).

Results are much better in the area of data collection. A large majority of the countries specify their **sources of data collection**, and six explicitly link indicators to data sources. Tanzania outperforms the other countries with regard to the horizontal logic of the indicator matrix: for each indicator in the M&E plan, the source, frequency and institution responsible for the data collection are specified. The generally favourable results in the area of data collection are in sharp contrast to the scores on **methodology**, with only three countries (Mozambique, Uganda and Rwanda) providing adequate information on the methodologies that will be used in their evaluation activities. For example, Mozambique discusses the importance of triangulation and mentions its intention to use both qualitative and quantitative methodologies, in addition to participatory methods for local-level data collection (Republic of Mozambique, 2006). Of the three countries, Rwanda is the only one that discusses impact-evaluation methodologies and the use of randomised controlled trials and quasi-experimental designs in this respect (Republic of Rwanda, 2007). This general lack of attention to methodological accuracy, particularly with regard to issues of validity, has also been observed in meta-evaluations within aid agencies (see, for example, Bollen et al., 2005). This is not without danger: deficient rigour in evaluation may lead to sub-optimal development outcomes, because of erroneous decisions not to invest in worthwhile policies and projects (false negatives) or to continue to invest in projects that are not effective (false positives) (see Shadish et al., 2002).

3.3 Organisation

As mentioned previously, efforts in the area of M&E development tend to focus on the more technocratic dimensions. Backed by empirical evidence (Bamberger, 2010; Bedi et al, 2006; World Bank, 2005, 2007), acknowledgement is gradually increasing of the fact that the organisational and institutional dimensions of M&E are often more influential and difficult to organise. This is also evident from our review: most countries have understood the importance of identifying the organisational structure for the **co-ordination**, support and central **oversight**, at least on paper. Several countries (for example, Benin, Burkina Faso, Cameroon) explicitly refer to the need to strengthen co-ordination and oversight. On paper, differences between countries do not seem to be

substantial: none of them scores 'weak' or 'excellent', most are considered 'partially satisfactory', while five countries (Benin, Ethiopia, Ghana, Tanzania, Zambia) have scores of 'satisfactory'.

Interestingly, countries that have longer experience with PRSPs do not necessarily outperform the others. For example, co-ordination and oversight remain areas of weakness for M&E in Uganda (see also Canagarajah and van Diesen, 2006). Particularly striking is the continuous change in the establishment of the system, making it increasingly complex and rendering oversight and co-ordination even more challenging. In addition, continuous reforms make it difficult for matters to be thought through practically, let alone to be implemented, tried out and left to mature. A similar 'reformitis' is also visible in the case of Mozambique, where the responsibility for M&E activities at the central level has recently shifted from one directorate to two others within the Ministry of Planning, thus scattering the M&E landscape even further (Republic of Mozambique, 2006: 158).⁴ Despite the emerging general acknowledgement of the vital importance of co-ordination amongst the various actors (for example, line ministries, statistical offices, decentralised entities) involved in various stages of M&E, in practice the effective organisation of co-ordination and oversight is clearly not a stroke-of-the-pen exercise. As highlighted by Bamberger (1991), who observed a similar pattern of ever-changing institutional M&E arrangements in South Asia, this situation is largely related to competition amongst agencies to control M&E and to the fear that units will become too powerful.

With regard to the identification of **the role of the statistical office in the M&E system**, our review indicates a considerable level of variation in the way this matter is addressed in various PRSPs. For example, Sierra Leone devotes major attention to the statistical office in its chapter on M&E, and it even includes a separate chapter on how to mainstream statistics into the PRSP. In order to improve the co-ordination of data-collection processes and to ensure a more timely data supply, a Memorandum of Understanding has been established between Sierra Leone Statistics and the line ministries, including a plan to install a statistical office in all ministries (Republic of Sierra Leone, 2009: 139). In sharp contrast are some other cases (for example, Burkina Faso, the Gambia, Malawi, Mauritania), which define the role of the statistical office only in broad terms, while failing to address the integration of the statistical office into the M&E system.

Most of the countries acknowledge the importance of sectoral M&E, as well as the need for stronger **horizontal linkage** amongst central and sector-specific M&E arrangements. Sector ministries are the main sources of information at the sectoral level and report on progress within their sectors. In most countries, sector-specific M&E arrangements either exist or are being set up, particularly in sectors where Sector-Wide Approaches (SWAs) are used (mostly education and health). Less attention is paid to the way in which various sector-specific M&E arrangements co-ordinate with each other and with the central M&E unit and the statistical office. This horizontal integration is particularly likely to have important payoffs in terms of the analytical

4. Interestingly, Mozambique and Uganda are also the two countries whose score on indicator 11 (quality of results-oriented frameworks) has decreased from the 2008 to the 2011 Paris Declaration survey (OECD, 2011: 138).

quality of evaluative exercises. In fact, it might help to address the problem of the ‘missing middle’,⁵ which seriously affects the ‘evaluability’ of PRSP-related interventions. In most countries (and often under the auspices of donors), there is a bias towards data collection at the level of inputs (PFM) and impact (MDGs) and less is known about the middle ground between the two extremes of the causal chain. Sector-management information systems might offer an important means of providing information on this ‘missing’ middle of activities, outputs and outcomes (Booth and Lucas, 2002; Lucas et al., 2004). Stronger M&E at the sector level, and particularly improved co-ordination amongst sector and central-level ministries, might thus enhance the identification of full causal chains, while stimulating the cross-reading of various sectoral and national-level survey data.

The lack of coherence between the central PRSP level and the sector level goes well beyond the domain of monitoring and evaluation, and it is often institutionally and politically difficult to address. It necessitates a more coercive mandate at the central level, in addition to more discipline on the side of the line ministries (for example, regarding the use of uniform formats and timetables). Notwithstanding the challenges encountered, some countries are undertaking efforts to boost horizontal integration. Rwanda, for instance, has located PRSP evaluation focal points in all sector ministries. In addition, a user-friendly format for reporting monitoring results has been proposed for use across sectors, and there are plans to achieve further harmonisation amongst the Management Information Systems that are in place within the various ministries (Republic of Rwanda, 2007: 154).

The establishment of decentralised M&E and **vertical integration** is even more problematic than horizontal linkages, because of the generally lower capacity at the local level. Having a functional decentralised M&E system is still far from a reality in the large majority of the countries. In such countries as Tanzania, Uganda and Zambia, decentralised or local-level monitoring units are in place on paper, although they are not (or not yet) functional, and their role and composition are not always clearly identified. In many cases, local entities are considered to be mere outposts for central-level data collection, despite the obvious need for a reverse information stream and for analysis and feedback into local decision-making.

One of the better-practice cases in terms of both **horizontal and vertical integration** is Ghana. The M&E plan of Ghana’s GPRS II provides information about the institutional framework for sector-specific and decentralised M&E (Republic of Ghana, 2008: 9-15). At the district level, M&E committees are feeding information to the regional-level committee, which links it back to the national committee. Detailed information is available on the composition of the various committees at the district and regional level, and the Annual Progress Report devotes specific attention to M&E at the district level. The result of integration efforts amongst the sectoral and central levels is also reflected in the GPRSP II indicator list: central national core indicators are specified at the level of outcome and impact, and sector indicators are specified at the level of inputs and outputs. This level of co-ordination cannot be achieved overnight; it

5. See Booth and Lucas (2002) for a discussion of the ‘missing middle’.

is the result of an incremental process that began with the elaboration of sectoral and district M&E guidelines in 2006.

One issue that remains particularly troublesome is the **link between donor project M&E and national M&E mechanisms**. In 18 of the 20 countries, there is hardly any co-ordination at this level. Despite the many emerging efforts to co-ordinate information on different aid inflows, few efforts are being made in the later phases of monitoring and evaluation. Although the lack of information on this issue in PRSPs could be due to the fact that such information is largely the responsibility of the donor, its relevance for national M&E systems is obvious. Niger is one of the exceptional cases involving the *ex-ante* assessment of donor-initiated projects and programmes in order to ensure that they fit within the broader objectives of the country (Republic of Niger, 2007). No further information is provided, however, on how this is to be accomplished, and the extent to which the M&E activities of the 'accepted' projects will be harmonised with national M&E arrangements is not clear. In addition to the enormous transaction-cost burdens generated by the insufficient harmonisation of project-specific and national M&E procedures, the reluctance of donors to adopt systematic procedures for communicating the findings of project evaluation to partner countries creates a 'learning' deficit.

Despite these weaknesses, the monitoring and, in particular, evaluation activities of donor projects remain important, as information from projects can provide interesting reality-checks for both policy-makers and budget-support donors, who are increasingly tending to focus on the aggregate 'policy' picture. Donors have a role to play, particularly with regard to impact evaluations of pilot projects. Given the general tendency towards underinvestment in this type of evaluation because of their public-good characteristics, joint donor involvement in terms of financial and technical capacity might be one way to lower this obstacle (see also Center for Global Development, 2006). To date, however, impact evaluations are often selected in an ad-hoc manner, depending largely upon the interests of donors (Bamberger, 2010) and often providing feedback only to the same projects or their donor headquarters (OECD/DAC, 2010). In order to exploit its full potential for development, donor project evaluation should be institutionalised within the national evaluation plan, which should entail the joint identification of evaluation needs and strategies in order to transfer information from the evaluations to national actors. This will also necessitate changes in the organisation of evaluation at the level of aid-agency headquarters and, more specifically, in the relationships between the field and headquarters (see also Liverani and Lundgren, 2007).

3.4 Capacity-building

In general, the diagnosis of **capacity problems** and the related **capacity-building plans** are amongst the best developed aspects of the M&E systems under study. Similar to our 2005 review, the focus remains predominantly on human-resources capacity-building in the area of data collection and statistics, although organisational issues are gradually moving into the foreground. Two countries, Benin and Mali, have scores of 'excellent' on this point. For example, Benin identifies challenges in the area of the institutional as

well as the incentive structure. Mali, which was also considered the most frank in recognising capacity problems in the 2005 review, now proposes a capacity-building plan that also includes M&E institutional development. The translation of these well developed capacity-building plans into practice presents an ongoing challenge. Although there are initiatives that focus on the elaboration and implementation of National Strategies for the Development of Statistics (NSDS; see www.paris21.org/NSDS), including the Partnership in Statistics for Development in the 21st Century (PARIS21), comparable initiatives for assisting countries in elaborating comprehensive M&E plans/systems have yet to be developed. Donor agencies, many of which have added capacity-building to their mandates, are often keen to provide support in this area, but are also faced with the same implementation problem. In addition, the lack of co-ordination between the various donors can lead to situations in which one aspect of the M&E system is being developed without considering other existing capacity-building initiatives, thereby resulting in loopholes and duplications in the system.

Practices that might be particularly valuable for evaluation capacity-building and which have largely remained unexplored in this setting are formative meta-evaluation (Stufflebeam, 1974) and twinning, which involves establishing linkages between institutions with similar functions but which are at different levels of development (Jones and Blunt, 1999).

3.5 Participation of non-governmental actors

The presence and functioning of M&E supply and demand outside national governmental structures is considered a key to the success of the new aid paradigm. Parliaments, civil-society organisations (both national and international), research institutes, national evaluation societies, donors and other actors may produce independent evidence and analysis about the implementation and impact of service-delivery and policy processes. They may also hold governments accountable, requesting reliable information and objective assessments of outcomes. Practice in this area, however, does not always develop parallel to the discourse.

Our review shows that the participation of non-governmental actors in M&E is generally low, particularly with regard to **Parliament**. Most of the countries reviewed (13 out of 20) do not even discuss the role of Parliament in M&E, and only four countries (Ghana, Malawi, Sierra Leone and Zambia) have scores of ‘satisfactory’ in this respect. For example, Ghana’s PRSP explicitly recognises the parliamentary oversight role, various parliamentary committees play an important role in M&E and APRs are submitted to the Parliament and are discussed during a workshop with the main stakeholders (Republic of Ghana, 2005). Sierra Leone’s PRSP includes activities intended to enhance Parliament’s research capacities, with the aim of reinforcing its capacity for oversight and scrutiny (Republic of Sierra Leone, 2009: 134).

Although more attention tends to be devoted to the role of **civil society** in M&E, procedures or institutional arrangements focused on the effective organisation of CSO involvement are often not specified. If anything, their participation is still more ad-hoc than institutionalised, and involves more consultation than real participation. Several

good practices are emerging in the case of Ghana, Mali, Mozambique and Uganda. In addition to CSOs, research and academic institutions are particularly involved in district-level M&E, using such methodologies as participatory poverty assessments and service-delivery surveys, which try to capture the extent to which beneficiaries are satisfied with the services. A well-known case is the Ugandan Debt Network, which is strongly involved in community-based monitoring and which co-ordinates the monitoring of PRSP-funded activities by CSOs at the district level. While CSOs also participate in the national-level technical meetings (Government of Uganda, 2007: 76), the 2006 APR acknowledges that information from the CSOs' monitoring activities is 'not fully channelled into mainstream government use in policy making' (ibid.: 77).

In many of the countries under study, **donor** agencies are the main sponsors of M&E, exerting major influence on the identification of M&E priorities and methodologies (see also Bamberger, 2000:96). With a few notable exceptions (for example, Mali, Sierra Leone, Guinea), the role of donors goes largely unnoticed in PRSP documents. It is not that the role and position of donors in this respect is without discussion. Some authors (for example, Gould and Ojanen, 2005: 63-4) argue that external pressure for donor accountability tends to crowd-out internal mechanisms of democratic accountability. Others (for example, Booth, 2005: 5) stress that it is 'unrealistic and unhistorical to believe that donors only have to back off from demanding accountability for aid funds in order for domestic accountability to flower in its place'. If anything, it is reasonable to question the independence of donors and their incentives to engage in proper evaluative exercises in the context of budget support, in which they align themselves with the policies and systems of particular countries and where the issue of attribution is difficult to address (see also De Renzio, 2006).

Another issue that has thus far remained underexplored is the importance of networking amongst various non-governmental actors in their efforts to forge domestic accountability (see also Santiso, 2008; Wang and Rakner, 2005). For example, research institutes and national evaluation societies could be valuable in filling the gaps left by inadequate CSO and parliamentary analytical research capacity. CSOs could have a comparative advantage in terms of accessibility to field data, while parliamentarians could have more leverage within the political system. For their part, donors might be able to broaden the space for national non-governmental actors to manoeuvre, particularly in cases where political-opportunity structures are closed. In fact, by merely using data from independent sources, donors may create incentives for data collection. Moreover, cross-reading and triangulation of data from various sources are also essential for the sake of the general validity of findings.

3.6 Use

Results are mixed with regard to the effective **use of M&E in Annual Progress Reports** (APRs). In the large majority (15 of 20) of the cases, data are used, evolutions in indicators are described, the achievement of targets is discussed and problems with implementation are highlighted. In contrast, the reports lack analytical depth, in the sense that results are compared with targets, leaving the reasons for the discrepancies unveiled. The lenient analytical quality is not particularly surprising; it is related to the emphasis on monitoring at the expense of evaluation (Item 2), the absence of causal

chains at the time of policy-making and planning (Item 9) and a failure to deliver M&E results on time (see also Thomas, 2010: 546). In countries where the APRs are also being used outside the PRSP context, it is likely that demand for, and incentives to produce, better quality M&E outputs will increase. Some evidence of this is visible in Ghana, where APRs are also serving as accountability instruments for other international agreements, including the Millennium Development Goals, the African Peer Review Mechanism, the Multi-Donor Budget Support (MDBS)/Poverty Reduction Support Credits (PRSC) triggers and targets (Republic of Ghana, 2008). Besides their 'external' use, the reports are also widely disseminated inside the country, and are submitted to Parliament and discussed in workshops.

These findings for Ghana are in contrast to most of the other countries under review. Although most PRSPs include a number of blunt generalisations about their intention to use M&E findings, the available information provides no indication that M&E outputs are effectively being **used for internal accountability and learning needs**. The following points could partially explain the apparent lack of use of M&E outputs for internal accountability and learning. First, as mentioned earlier, use is obviously constrained by several factors, including analytical shallowness and the timely availability of findings. Niger, for example, explicitly mentions that 'the annual implementation reports cannot serve as decision-making instruments for the political authorities because of the inadequate information and the poor analysis' (Republic of Niger, 2007: 354). Similar remarks about the difficulty of using M&E outputs for decision-making can be found in the APR and JSAN of other countries such as Mali and Uganda. Next, although lack of methodological rigour is an important stumbling block for the use of M&E in decision-making, conversely having high-quality M&E outputs does not guarantee their use. Other relevant factors (see Cousins and Leithwood, 1987, Fleischer and Christie, 2009) are also mediating the relationship between M&E outputs and their use for decision/policy-making. Mackay's (2006: 5) definition of an institutionalised M&E system, for example, implies that a substantive M&E demand combined with the right incentive structure can stimulate the utilisation of M&E outputs, and contribute to the sustainability of the system. Finally, the complexity of evaluation use and influence on policy and the multiple ways in which such influence can occur have been pointed out by several authors (Weiss, 1999; Mark and Henry, 2004). The type of use that is assessed with our checklist is mainly direct, instrumental use of M&E results in budgeting and decision/policy-making as reported in APRs. Research on evaluation use and influence, however, suggests that instrumental use is not very common and that evaluations are more likely to contribute in terms of bringing new information, ideas and perspectives into the policy arena (Weiss 1999: 471), which is also called enlightenment or conceptual use. Field work, rather than a desk-based review, is required to study these types of evaluation use and influence.

4 Conclusion

The changing development architecture reconfirms the importance of M&E, while making weighty demands for reform with regard to M&E for both donors and recipients. From the perspective of aid agencies, the principles that pose the greatest

challenge (co-ordination and stakeholder participation) have moved to the top of the agenda. At a time when donors are facing pressure from the increasing accountability demands of their own constituencies, they are being asked to refrain from using their own parallel M&E systems and to align them as much as possible with those of the recipient countries, which they often consider fragmentary at best. For their part, recipients are being urged to upgrade their M&E arrangements. While most donor agencies have added M&E capacity development to their mandates, strategies on how to translate discourse into practice are generally lacking thus far. This article aims to contribute to this point.

Proceeding from the observation that small incremental changes to the existing situation might be more feasible and workable than radical changes, the first step in any capacity-development initiative should be diagnosis. In order to move beyond the fairly restrictive scope of the CDF 'results focus' indicator, thereby increasing construct validity, we have elaborated a checklist of 23 items that capture dimensions of M&E policy, indicators, data collection and methodology, organisation, capacity-building, participation of actors outside government and use. This checklist has subsequently been used to assess the M&E arrangements that have been elaborated by the 20 SSA countries that had prepared at least a second Poverty Reduction Strategy Paper as of May 2010.

Our findings generally support earlier evidence that indicators and statistics are often amongst the most developed components of recipient M&E systems. The large majority of the countries under review have identified indicators and targets, as well as instruments for data collection. The bias towards 'statistics' is also obvious from a country's evaluation policies: while most countries do not (or not yet) have comprehensive evaluation plans outlining what to evaluate, how, why and for whom, they do have some type of strategy in place for the development of statistics. Our findings also suggest that performance monitoring may be crowding out evaluation, a phenomenon that is also observed within aid agencies and reinforced by the general shift towards result-based management in development co-operation. The neglect of evaluation is not surprising. Evaluation is methodologically challenging, particularly in settings where policies and programmes have not been conceptualised with a programme theory in mind and where indicators are thus not specified in terms of inputs, outputs, outcomes and impact, let alone linked together in some form of logic chain. Elaborating full causal chains also necessitates co-ordination amongst central-level ministries, most of which are dealing with the two extremes (inputs and impact), and sectoral ministries, which are largely focusing on outputs and outcomes (the 'missing middle'). If anything, co-ordination amongst the various actors and institutions involved in an M&E system is not a stroke-of-the-pen exercise. While the need for oversight and co-ordination is increasingly acknowledged, practice does not often develop parallel to discourse, as it touches upon the underlying institutional processes, which are more difficult to address.

Particularly when it focuses on 'aggregates', monitoring is also politically safer than the more analytical evaluative undertakings, which allow the distinction between 'doing things right' and 'doing the right things'. In a similar vein, the silence of M&E chapters on the key M&E principles of 'autonomy' and 'impartiality' is understandable, as is the meagre attention paid to independent sources of M&E demand and supply.

While these are vital to the realisation of accountability, they may also jeopardise the power positions of the government. Although it is methodologically and politically comfortable to downplay evaluation, doing so could lead to the production of outputs that are analytically shallow and hardly ever used to satisfy the need for accountability and learning. This holds for internal usage (in the context of national and local planning and budgeting) as well as for external use (in the context of budget support). It is highly unlikely that the lenient quality of M&E outputs will be able to persuade donors, and particularly their home constituencies, to give the M&E systems of recipient countries the benefit of the doubt. If anything, it will not stimulate the dismantling of the donors' own parallel M&E systems. Our findings demonstrate that issues involving the linkage between donor and national M&E persistently stand amongst the least developed aspects. Isolated donor M&E places a heavy burden on recipient M&E arrangements and creates a learning deficit, as neither findings nor methodologies are disseminated beyond the specific donor project setting. Donor project monitoring and particularly evaluation have not become superfluous; on the contrary, it is in the area of evaluation of the impact of pilot interventions that the joint financial and technical involvement of donors is particularly likely to lower the constraints related to the public-good character of impact evaluations. In order to maximise feedback and learning for all stakeholders involved at the national level, there is a particularly urgent need for the institutionalisation of impact evaluation in the evaluation policies of recipient countries. Moreover, in order to stimulate cross-country learning, donors may also jointly invest in meta-evaluative and synthesis exercises.

Our own cross-country exploratory exercise has highlighted a number of patterns in the current status and development of recipient M&E arrangements, with the aim of informing joint efforts to strengthen these systems. Throughout this article, we have highlighted the need to complement the present desk-based study with field research in order to obtain more insight into the subtleties of the existing arrangements/system and study more closely issues related to the use and influence of M&E. Political-economy analysis could be particularly useful here to analyse the underlying institutional issues and informal 'power' processes at play which are often more influential than M&E technicalities. In combination with a comparative case-study approach, it also allows studying more closely how and why different sub-sections of national M&E systems develop differently. Finally, an issue which has thus far remained largely unexplored is the role national evaluation societies could play in national M&E development. Studying how national evaluation societies, which exist in most of the countries under study, could create a networking forum among the M&E supply and demand side and the impact this might have on the use/influence of M&E is theoretically interesting as well as policy-relevant.

first submitted April 2011
final revision accepted February 2012

References

- Bamberger, Michael (2010) 'Institutionalising Impact Evaluation. A key element in strengthening country-led monitoring and evaluation systems' in M. Segone (ed), *From Policies to Results. Developing capacities for country monitoring and evaluation systems*. Geneva: UNICEF.
- Bamberger, Michael (2000) 'The Evaluation of International Development Programs: A view from the front', *American Journal of Evaluation* 21(1): 95-102.
- Bamberger, Michael (1991) 'The Politics of Evaluation in Developing Countries', *Evaluation and Program Planning* 14: 325-39.
- Bedi, Tara; Coudouel, Aline; Cox, Marcus; Goldstein, Markus and Thornton, Nigel (2006) *Beyond the Numbers. Understanding the Institutions for Monitoring Poverty Reduction Strategies*. Washington, DC: World Bank.
- Bietenhader, Daniel and Bergmann, Andreas (2010) 'Principles for Sequencing Public Financial Reforms in Developing Countries', *International Public Management Review* 11 (1): 52-65.
- Bollen, Kenneth, Paxton, Pamela and Morishima, Rumi (2005) 'Assessing International Evaluations. An Example from USAID's Democracy and Governance Program', *American Journal of Evaluation* 26 (2): 189-203.
- Booth, David (2005) *Missing Links in the Politics of Development: Learning from the PRSP experiment*. ODI Working Paper No. 256. London: Overseas Development Institute.
- Booth, David and Lucas, Henry (2002) *Good Practice in the Development of PRSP Indicators and Monitoring Systems*. ODI Working Paper No. 172. London: Overseas Development Institute.
- Canagarajah, Sudharshan and van Diesen, Arthur (2006) 'The Poverty Reduction Strategy Approach Six Years On', *Development Policy Review* 24 (6): 647-67.
- Center for Global Development (2006) *When Will We Ever Learn? Improving lives through impact evaluation*. Washington, DC: Center for Global Development.
- Cousins, Bradley and Leithwood, Kenneth (1986) 'Current Empirical Research on Evaluation Utilization', *Review of Educational Research* 56 (3): 331-64.
- Davies, Huw, Nutley, Sandra and Smith, Peter (1999) 'Introducing Evidence-based Policy and Practice in Public Services' in H. Davies, S. Nutley and P. Smith (eds), *What Works? Evidence-based Policy and Practice in Public Services*. Bristol: The Policy Press.
- De Renzio, Paolo (2006) 'Aid, Budgets and Accountability: A Survey Article', *Development Policy Review* 24 (6) : 627-45.
- Donaldson, Stewart; Christie, Christina and Mark, Melvin (2008) *What Counts as Credible Evidence in Applied Research and Evaluation Practice*. London: Sage Publications.
- Fleischer, Dreolin and Christie, Christina (2009) 'Evaluation Use. Results from a Survey of U.S. American Evaluation Association Members', *American Journal of Evaluation* 30 (2): 158-75.
- Gordillo, Gustavo and Andersson, Krister (2004) 'From Policy Lessons to Policy Actions: Motivation to take evaluation seriously', *Public Administration and Development* 24 (4): 305-20.

- Gould, Jeremy and Ojanen, Julia (2005) 'Tanzania: Merging in the Circle' in J. Gould (ed), *The New Conditionality. The Politics of Poverty Reduction Strategies*. London: Zed Books.
- Government of Uganda (2007) *The 2006 Annual PEAP Implementation Review: A Synopsis*. Kampala: Government of Uganda, Office of the Prime Minister, November.
- 3rd High Level Forum on Aid Effectiveness (2008) *Accra Agenda for Action*. Accra: 3rd High Level Forum on Aid Effectiveness.
- Holvoet, Nathalie and Inberg, Liesbeth (2012) *Sector Monitoring and Evaluation Systems in the Context of Changing Aid Modalities: The case of Uganda's Health Sector*. Antwerp: Institute of Development Policy and Management, University of Antwerp.
- Holvoet, Nathalie and Inberg, Liesbeth (2011) *Stocktaking and Assessing M&E Arrangements in Rwanda's Health Sector: Evidence from Desk and Field Study*. Study in the context of the formulation of Belgian Sector Budget Support to Rwanda's Health Sector. Antwerp: Institute of Development Policy and Management, University of Antwerp.
- Holvoet, Nathalie and Renard, Robrecht (2007) 'Monitoring and Evaluation under the PRSP: Solid Rock or Quicksand?', *Evaluation and Program Planning* 30: 66-81.
- Holvoet, Nathalie and Rombouts, Heidi (2008) 'The Challenge of Monitoring and Evaluation under the New Aid Modalities: Experiences from Rwanda', *Journal of Modern African Studies* 46 (4): 577-602.
- IDA and IMF (2008) *Annual Poverty Eradication Plan (PEAP) Implementation Review Joint Staff Advisory Note*. Washington, DC: International Development Association and International Monetary Fund, 22 December.
- IMF and IDA (2006) *Republic of Senegal, Joint Staff Advisory Note on the Second Poverty Reduction Strategy Paper*. Washington DC: International Monetary Fund and International Development Association, December.
- Isham, Jonathan, Narayan, Deepa and Pritchett, Lant (1995) 'Does Participation Improve Performance? Establishing Causality with Subjective Data', *The World Bank Economic Review* 9 (2): 175-200.
- Jones, Merrick L. and Blunt, Peter (1999) "'Twinning" as a Method of Sustainable Institutional Capacity Building', *Public Administration and Development* 19 (4): 381-402.
- Kusek, Jody Z. and Rist, Ray C. (2004) *Ten Steps to a Results-Based Monitoring and Evaluation System. A Handbook for Development Practitioners*. Washington, DC: World Bank.
- Liverani, Andrea and Lundgren, Hans E. (2007) 'Evaluation Systems in Development Aid Agencies. An Analysis of DAC Peer Reviews 1996-2004', *Evaluation* 13 (2): 241-56.
- Lucas, Henry; Evans, David; Pasteur, Katherine and Lloyd, Robert (2004) *Research on the Current State of PRS Monitoring Systems*. IDS Discussion Paper No. 382. Brighton: Institute of Development Studies at the University of Sussex.

- Mackay, Keith (1999) *Evaluation Capacity Development: A Diagnostic Guide and Action Framework*. ECD Working Paper Series No.6. Washington, DC: World Bank.
- Mackay, Keith (2006) *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management*, ECD Working Paper Series No. 15. Washington, DC: World Bank.
- Mark, Melvin and Henry, Gary (2004) 'The Mechanisms and Outcomes of Evaluation Influence', *Evaluation* 10 (1): 35-57.
- Ministry of Economy and Development of Burkina Faso (2004) *Poverty Reduction Strategy Paper*. Ouagadagou: Ministry of Economy and Development, July.
- Nairobi M&E Network, African Evaluation Association Secretariat, Réseau Nigérien de Suivi et Evaluation, Cape Verde Evaluation Network, Réseau Malagache de Suivi et Evaluation, Comores Evaluation Network, Eritrean Evaluation Network, Malawi M&E Network, Réseau National de Chercheurs et Evaluateurs de Burundi, Rwanda Evaluation Network, UNICEF Eastern and Southern Africa Region M&E Network (2002) 'The African Evaluation Guidelines: 2002. A Checklist to assist in planning evaluations, negotiating contracts, reviewing progress and ensuring adequate completion of an evaluation', *Evaluation and Program Planning* 25 (4): 481-92.
- North, Douglass C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- OECD (2011) *Aid Effectiveness 2005-10. Progress in Implementing the Paris Declaration*. Paris: OECD.
- OECD/DAC (2010) *Development Evaluation Resources and Systems: A study of network members*. Paris: OECD/DAC.
- OECD/DAC (2008) *2008 Survey on Monitoring the Paris Declaration. Effective Aid by 2010. What it will take?* Paris: OECD/DAC.
- OECD/DAC (2007) *2006 Survey on Monitoring the Paris Declaration*. Paris: OECD/DAC.
- OECD/DAC (2006) *Evaluation Systems and Use. A working tool for peer reviews and assessments*. Paris: OECD/DAC.
- OECD/DAC (2005) *Paris Declaration on Aid Effectiveness*. Paris: OECD/DAC.
- OECD/DAC (1991) *OECD/DAC Principles for the Evaluation of Development Assistance*. Paris: OECD/DAC.
- Pollitt, Christopher (2001) 'Integrating Financial Management and Performance Management', *OECD Journal on Budgeting* 1(2): 7-37.
- Pritchett, Lance (2002) 'It Pays to be Ignorant: A Simple Political Economy of Rigorous Program Evaluation', *Policy Reform* 5 (4): 251-69.
- Ravallion, Martin (2006) 'Looking Beyond Averages in the Trade and Poverty Debate', *World Development* 34 (8): 1374-92.
- Republic of Ghana (2008) *National Monitoring and Evaluation Plan (2006-2009)*. Accra: Republic of Ghana.
- Republic of Ghana (2005) *Growth and Poverty Reduction Strategy (GPRSII) (2006-2009)*. Accra: Republic of Ghana, November.
- Republic of Mozambique (2006) *Action Plan for the Reduction of Absolute Poverty 2006-2009 (PARPA II)*. Maputo: Republic of Mozambique, 2 May.

- Republic of Niger (2007) *Accelerated Development and Poverty Reduction Strategy 2008-2012*. Niamey: Republic of Niger, Prime Minister's Office, PRS Permanent Secretariat, August.
- Republic of Rwanda (2007) *Economic Development and Poverty Reduction Strategy, 2008-2012*. Kigali: Republic of Rwanda (September 2007).
- Republic of Sierra Leone (2009) *An Agenda for Change: Second Poverty Reduction Strategy 2008-2012*. Freetown: Republic of Sierra Leone.
- Santiso, Carlos (2008) 'Eyes Wide Shut? Reforming and Defusing Checks and Balances in Argentina', *Public Administration and Development* 28: 67-84.
- Shadish, William R., Cook, Thomas D. and Campbell, Donald T. (2002) *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA and New York: Houghton Mifflin.
- Stufflebeam, Daniel L. (1974) *Meta-evaluation*. Occasional Paper Series 3. Kalamazoo, MI: Western Michigan University Evaluation Center.
- Thomas, Vinod (2010) 'Evaluation Systems, Ethics and Development Evaluation', *American Journal of Evaluation* 31 (4): 540-8.
- Valadez, Joseph, and Bamberger, Michael (1994) 'Organizational and Management Issues in Programme Evaluation' in J. Valadez and M. Bamberger (eds), *World Bank: Monitoring and Evaluating Social Programmes in Developing Countries: A Handbook for Policymakers, Managers and Researchers*. Washington, DC: World Bank.
- Van de Walle, Nicolas (2005) *Overcoming Stagnation in Aid-Dependent Countries*. Washington, DC: Center for Global Development.
- Wang, Vibeke and Rakner, Lise (2005) *The Accountability Function of Supreme Audit Institutions in Malawi, Uganda and Tanzania*. Bergen: Chr. Michelsen Institute.
- Weiss, Carol (1999) 'The Interface Between Evaluation and Public Policy', *Evaluation* 5 (4): 468-86.
- Wood, Bernard; Kabell, Dorte; Muwanga, Nansozi and Sagasti, Francisco (2008) *Synthesis Report on the First Phase of the Evaluation of the Implementation of the Paris Declaration*. Copenhagen: Ministry of Foreign Affairs.
- World Bank (2007) *Results-Based National Development Strategies Assessment and Challenges Ahead*. Washington, DC: World Bank.
- World Bank (2005) *Getting Serious About Meeting the Millennium Development Goals. A Comprehensive Development Framework Progress Report*. Washington, DC: World Bank.
- World Bank (2003) *A User's Guide to Poverty and Social Impact Analysis*. Washington, DC: World Bank.

Annex 1: Diagnostic checklist

Items	Question
I. Policy	
1. Evaluation plan	Is there a comprehensive evaluation plan, indicating what to evaluate, why, how, for whom?
2. M versus E	Is the difference and the relationship between M and E clearly spelled out?
3. Autonomy and impartiality	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, integration?
5. Alignment with planning and budgeting	Is there integration of M&E results in planning and budgeting?
II. Indicators, Data Collection and 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)?
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?
III. Organisation	
12. Co-ordination and oversight	Is there an appropriate institutional structure for co-ordination, 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 a central unit?
15. Decentralised levels	Are there M&E units at decentralised levels and are these properly relayed to a central unit?

-
- | | |
|------------------------|--|
| 16. Link with projects | Is there any effort to relay/co-ordinate with donor M&E mechanisms for projects? |
|------------------------|--|

IV. Capacity-building

- | | |
|----------------------------|--|
| 17. Capacity diagnosis | Are current capacity strengths and weaknesses identified? |
| 18. Capacity-building plan | Are there plans for remediation? Do these include training, appropriate salaries, etc. |

V. Participation of non-governmental actors

- | | |
|-------------------|---|
| 19. Parliament | Is the role of Parliament properly recognised, and is there alignment with parliamentary control and oversight procedures? |
| 20. Civil society | Is the role of civil society recognised? 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 recognised? Are there clear procedures for participation of donors? |

VI. Use

- | | |
|-------------------------------|--|
| 22 In Annual Progress Reports | Is there a presentation of relevant M&E results? Are results compared with targets? Is there an analysis of discrepancies? |
| 23 Within country | Are M&E outputs (e.g. APR) also used for internal purposes? E.g. within national policy-making and/or policy-influencing and advocacy? |
-