NETwork for Research and Evaluation on Education and Development

NETREED Conference 2005

Evaluation of education—on whose terms?

Beitostølen, Norway, 5th-7th December 2005

Evaluating education projects and programmes: on whose terms and to what effect?

Malcolm Mercer[1]

Consultant in education planning, management and finance: Powys, UK

Abstract: Who evaluates? Who benefits? This paper examines the practice of evaluating education projects and programmes financed by external funding and technical assistance (or donor) agencies. It takes a critical look at some of the various rationales behind, and processes and effects of, such evaluation. Illustrative examples are drawn mainly from the author's experience in monitoring and evaluating education projects and programmes funded by the European Commission, the UK Department for International Development, the Asian Development Bank and other agencies in Africa, the Caribbean, and Central and South Asia. The paper concludes by summarising lessons learned and expressing the need for follow-up research.

INTRODUCTION

Monitoring and evaluation have in theory been integral parts of the process of planning and implementing education sector activities. In recent years, theory has increasingly been translated into practice and evaluation takes place at a number of levels—the classroom level, the community level, the institutional level, the individual project level, the programme or multiple project level, the programme or sub-sectoral level, and the level of the education sector as a whole.

External funding and technical assistance ("donor") agencies in particular require that the programmes or projects which they finance are subject to regular monitoring, to mid-term evaluation and to final evaluation near, upon or after completion. This paper, drawing upon the author's experience in Africa, the Caribbean and Central and South Asia, examines the various types of evaluation, the rationales behind evaluation, and the processes and effects of evaluating education sector activities at project, programme and sector levels.

WHO EVALUATES, WHY, FOR WHOM AND HOW?

Riddell (1999:387) summarised the focus and interests of actors in educational evaluation in a table, reproduced here in a slightly modified form as Table 1. In answer to the question 'Who evaluates?' it is clear that we *all*—from students, parents and teachers through education administrators at different levels to donor agencies—evaluate education activities and processes at one time or another. However, our focus differs according to the level in which we are operating.

We are also concerned with different types of evaluation and reasons for them. Broadly speaking, we monitor activities or progress, generally on a regular basis, to find out and make a record of what is going on, whether plans are being adhered to and whether objectives are being achieved or not. Advancing further, we carry out *formative* evaluation during the process of a

---

[1]   Correspondence via e-mail to MercerMM@aol.com.

project or programme in order to determine what is successful and what is not and, most importantly, why, i.e. in order to determine causal inferences. We also examine the effect or impact of activities on direct beneficiaries (e.g. learners) and indirect beneficiaries (e.g. members of the community in which the learners live) and we consider or recommend ways in which greater success can be achieved for the various actual or potential beneficiaries. Towards the end of a project or programme or, as is more frequently the case, some time after a project or programme has been completed, we carry out *summative* evaluation for accountability to determine how efficiently resources have been used, the extent to which the programme has been a success, what impact it has had on the wider population and environment and whether it is worth replicating in the same area or elsewhere.

Table 1: Focus and interests of actors in educational evaluation

Actors/purposes	Focus	Formative evaluation	Summative evaluation
Donor agency	Whole programme, possibly relative to other countries	Dependent on programme timing and renewal. Lessons learned? Factors/contexts of success/failure?	Programme success? Impact? Worth funding? Worth replicating?
National Ministry of Education	Whole programme relative to other programmes	What should be changed for more success—nationally and for particular areas/groups?	Worth funding? Worth participating? Worth replicating?
Project manager	Whole programme	Dependent on programme timing and renewal. Lessons learned? Factors/contexts of success/failure?	How do I look? Success *versus* mismanagement
Regional and district education administration	The region or district relative to others and the whole programme	What should be changed for more success—how do we account for differences across regions/districts?	How do we look? Success *versus* mismanagement
Head teacher	The school, relative to others	What influence has the programme had on particular types of student or teacher? What lessons for further development?	Did it make any difference (given the context)? Worth participating?
Teacher	The class/es, relative to others	How was I influenced? How were particular students influenced? What lessons for further development?	Did it make any difference? Worth participating?
Parents, students, communities	The child and his/her own prospects, the community, relative to others	How was the community affected? How are siblings likely to be affected?	Did s/he benefit? Did the community benefit? Worth participating?

Source: Riddell (1999:387)

The evaluation of education projects and programmes may be undertaken at the instigation, and for the benefit, of a number of clients. Firstly, it can be carried out as part of the research programme of an individual, a university department or a research institution. The results of this research may be available for wider circulation rather than only for a particular financing body. Secondly, an education project may be evaluated for the benefit of the contractor responsible for managing the project in order to give guidance to the contractor on best practice in future. Thirdly, the Ministry of Education or a country's major education service provider may request evaluation to take place. Such evaluation is often internal and carried out only by staff of the agency or institution managing a programme; and, to be of any value, it is best done in a corporate ethos that encourages openness, constructive criticism and the capacity to learn from their own and others' experience. Finally, donor agencies, frequently in cooperation with government agencies, carry out (external) formative and summative evaluation exercises as agreed in the programme terms of reference or implementation plan.

Broadly speaking, external evaluation can be carried out in five ways: (i) by an individual or individuals contracted specifically for the purpose[2]; (ii) by a team of international and national consultants[3]; (iii) by donor agency staff only[4]; (iv) by one external consultant or more (international and/or national) together with donor agency staff and programme staff (national and/or international)[5]; or (v) as in (iv) with the addition of staff of the client government.[6]

Although multilateral lending agencies such as the World Bank and the Asian Development Bank (ADB) put great emphasis on economic analysis as a starting point both for setting priorities and for evaluating past performance, at another level the most appropriate place to start the evaluation process is to examine the programme or project design. The Logical Framework or LogFrame has been used in project design for two decades and more and is now *mandatory* with most donor agencies.[7] If constructed well and used properly the LogFrame can be a valuable tool for monitoring and evaluation, but unfortunately it is still little understood by many programme or project designers. Although the logic of cause and effect from bottom to top is straightforward enough, the language used for each level of objectives can be confusing even for a native speaker of English, as can be seen in Table 2 which illustrates the different terms used for describing objectives and the relationship between them in the first column (out of four columns) commonly called the Narrative Summary or Design Summary. For example, where other agencies use the term *Goal*, ADB uses *Impact*; where others use the term *Purpose*, ADB uses *Outcome*; and where others use the term *Outputs*, the European Commission (EC) uses *Results*.[8]

Table 2: Logical Framework hierarchy of objectives or Narrative Summary

Narrative Summary / Design Summary (= hierarchy of objectives)		... Indicators
Goal / Impact [ADB] (= sector/programme/wider/development/overall objective)	*linked to **purpose** through the terms* impact, outcome *or* overall effect *on larger numbers of people* [EC]	
Purpose / Outcome [ADB] (= impact or project/immediate/specific objective)	*linked to **results** through the term* effectiveness [EC]	
Outputs / Results [EC] (= deliverables or terms of reference)	*linked to **activities** and **resources/means** through the term* efficiency [EC]	
Activities / Activities with milestones [ADB] (= key activity clusters)		Inputs/Resources/Means

---

[2]   As, for example, in Mercer *et al.* (2000) where three individuals were recruited (and paid for separately) by the European Commission (EC), the UK Department for International Development (DFID) and Ireland Aid to evaluate (as a team) three separate projects which each had the aim of supporting the design, preparation and management of Tanzania's Education Sector Development Programme (ESDP).

[3]   As, for example, in Mercer *et al.* (2002a) where a team of two international and three national consultants was contracted by the donor agency (the EC) to evaluate EC support to primary education in India, and in Mercer *et al.* (2001b, 2001c and 2002b) where a team of six international consultants contracted by the donor agency to evaluate EC support to the education sector in African, Caribbean and Pacific (ACP) countries was assisted by one national consultant in each of the five countries visited.

[4]   As, for example, in Asian Development Bank (2003) where two ADB staff members carried out a project completion review (PCR) after completion of the ADB-financed Education Sector Development Program (ESDP) in Mongolia. PCRs are generally carried out between 12 and 24 months after project completion.

[5]   As, for example, in Passingham *et al.* (1998) where one member of DFID staff, one external consultant and two representatives of the firm contracted to manage a primary education planning project in Sri Lanka, financed by DFID, carried out a mid-term (or Output to Purpose) review.

[6]   As, for example, in Mercer *et al.* (2001a) where one member of DFID staff, one external consultant, the (international) project manager and a ministry of education director carried out a mid-term review of a DFID-financed education management project in The Gambia.

[7]   For example, the USA, UK and GTZ have used the LogFrame in project design from at least 1985, the EC from 1993, the World Bank from 1997, and the ADB from 1997 with design and nomenclature amendments being made in 2005.

[8]   The Chambers English Dictionary (1990) offers the following (selected) meanings to some of the terms commonly used in LogFrames: (i) *Output* = (a) quantity produced or turned out; (b) data produced after processing by a computer; (ii) *Result* = (a) consequence; (b) outcome; (c) quantity obtained by calculation; (d) final score (in games); (iii) *Outcome* = (a) the issue; (b) consequence; (c) result; (iv) *Impact* = (a) the blow of a body in motion impinging on another body; (b) the impulse resulting from a new idea or theory; (c) strong effect or influence.

An outline of the LogFrame as a whole is given in Table 3. Here we can see the different terminology used for describing items to be included in the second, third and fourth columns, depending on the donor agency which is assisting in financing the programme or project. While the first column is a general summary of the objectives at different levels, the second column adds meaning to the narrative/design summary by elaborating upon the objectives in terms of *quantity* or how much is to be achieved, *quality* or what type is to be achieved, and *time* or by when the objective is to be achieved. Place, actors and cost can also be included in this column, although it must be remembered that the LogFrame is meant to be a framework only and not a full project description.

Table 3: Logical Framework outline

Narrative Summary or Design Summary	Objectively Verifiable Indicators (OVI) or Performance Indicators or Performance Targets	Means of Verification (MOV) or Sources of Verification or Data Sources or Monitoring Mechanisms or Reporting Mechanisms	(Important) Assumptions and Risks (events or activities outside the immediate control of the project management but which need to be taken into consideration as they may affect achievement of project objectives)
Goal	(quantity, quality & time)		
Purpose	(quantity, quality & time)		
Outputs	(quantity, quality & time)		
Activities	Inputs/resources/means		

Whereas the second column gives us details indicating where we are going or the extent to which project objectives are expected to be achieved, the third column tells us how we know that we have reached our destination or how we know the extent to which project objectives have been achieved and why. The means of verification should include details of the sources of information (such as key informant interviews, project records and reports, data collected by government or other agencies, and data collected through special studies or surveys[9]), how these sources are accessed (e.g. through an education management information system), who collects, reviews and audits the information and who accepts or agrees the result.

This is not the place to go into detailed discussion of the LogFrame. Suffice it to say that, for the purposes of evaluation, the second and third columns are crucial starting places. Great attention needs to be given to completing these columns both at the design stage and at intermediate stages of the project where amendments to the original design are likely to take place. Unfortunately this is seldom the case, particularly with the means of verification, and evaluation, especially summative evaluation, is thus made the more difficult.

As shown in Table 4, donor agencies evaluate programmes and projects under similar categories. The EC asks evaluators to examine aspects of a project in a strictly sequential order, beginning with (i) an assessment of the relevance of project design—both when the project was designed and at the time of evaluation—and the extent to which its stated objectives correctly addressed the identified problems or real needs. This is followed by an examination of (ii) how well the various activities transformed the available resources into the intended outputs—the project's efficiency; (iii) how far the project's results or outputs were used or their potential benefits were realised—the project's effectiveness in achieving the project purpose; (iv) the extent to which the benefits received by target beneficiaries had a wider overall effect on larger numbers of people—the project's impact or the relationship between achievement of the project purpose and achievement of the planned overall objectives; and (v) whether the positive outcomes of the project at purpose level are likely to continue after external funding ends—sustainability. The last element includes the evaluation of, for example, ownership of design and achievements,

---

[9]   Special studies or surveys, including base-line studies, need to be included in programme design and costed accordingly.

institutional capacity, socio-cultural factors, financial sustainability, technical/technological issues, cross-cutting issues such as gender equity, environmental impact and good governance, and over-arching issues such as poverty alleviation.

Table 4: What is evaluated by the EC and by the ADB?

EC	ADB
(i) Relevance (and design)	(i) Relevance (and design)
(ii) Efficiency (means/activities to results)	(ii) Efficacy (effectiveness) in achievement of purpose
(iii) Effectiveness (results to purpose)	(iii) Efficiency in achievement of outputs and purpose
(iv) Impact (purpose to overall objective)	(iv) Sustainability
(v) Sustainability	(v) Environmental, socio-cultural and other impacts

Source: EC (2001) and ADB (2003)

The ADB's list for evaluators is similar to that of the EC but in a slightly different order. Following an examination of project design and relevance, the evaluation begins near the top of the hierarchy of objectives and considers the effectiveness or, in old ADB terminology, *efficacy* in achieving the project purpose. The bulk of the evaluation considers efficiency in the achievement of outputs and purpose; this is mainly because the ADB is a bank and is concerned to a large extent with cost recovery and a country's ability to repay a loan, however concessional the terms of the loan may be. Sustainability and the project's environmental, socio-cultural and institutional development impact(s) are important elements of the evaluation, but only passing attention is given to the extent to which the project's overall objective or goal has been achieved. This is understandable since achievement of a project's goal is generally contingent upon the outcome of other projects and government's and other agencies' activities in the sector. Furthermore, it is often only possible to measure the extent to which the goal, and even the project purpose, has been achieved some years after project completion. However, in the overall assessment of the ESDP in Mongolia (ADB 2003:52), efficacy is given the largest weight (25%), compared with relevance, efficiency and sustainability (20% each) and institutional and other impacts (15%).

Evaluators are asked to give a performance rating both for different aspects of a project and for the project overall. Ratings differ according to the donor agency, as can be seen from Table 5.

Table 5: Examples of overall performance ratings for programmes, according to donor agency

ADB[10]	EC[11]	DFID[12]
≥ 20 = Highly successful	1 = Highly satisfactory (fully according to plan or better)	1 = likely to be completely achieved
16-19 = Successful	2 = Satisfactory (on balance according to plan, positive aspects outweighing negative aspects)	2 = likely to be largely achieved
11-15 = Partly successful	3 = Less than satisfactory (not sufficiently according to plan, taking into account the evolving context; a few positive aspects, but outweighed by negative aspects)	3 = likely to be partially achieved
≤ 10 = Unsuccessful	4 = Highly unsatisfactory (seriously deficient, very few or no positive aspects)	4 = only likely to be achieved to a very limited extent
		5 = unlikely to be realised
		x = too early to judge the extent of achievement

---

[10] In ADB (2005), performance rating points, with a possible total of 24, are awarded in five categories as follows:
  Relevance: highly relevant = 3 points; relevant = 2 points; partly relevant = 1 point; irrelevant = 0 points.
  Effectiveness: highly effective = 6 points; effective = 4 points; less effective = 2 points; ineffective = 0 points.
  Efficiency: highly efficient = 3 points; efficient = 2 points; less efficient = 1 point; inefficient = 0 points.
  Sustainability: most likely = 6 points; likely = 4 points; less likely = 2 points; unlikely = 0 points.
  Impact: high = 6 points; substantial = 4 points; modest = 2 points; negligible = 0 points.
[11] For mid-term and final evaluations, as in Mercer *et al.* (2002a).
[12] For Output to Purpose (mid-term) reviews, as in Mercer *et al.* (2001a).

ISSUES IN EVALUATING EDUCATION PROGRAMMES

There are several issues arising from the evaluation of externally funded education programmes and projects, but before discussing them in some detail it is worthwhile looking at the wider picture in the education sector as a whole.

Education sector analysis. In his examination of externally initiated, commissioned and supported studies of education in Africa, Samoff (1999) finds that they rarely touch upon the aid relationship itself. They do not attempt to analyse the nature and consequences of the increasing reliance on external assistance to support reform and even basic services in African education. Nor do they present us with empirical evidence on the relationship between aid and the quality of education in Africa. The reasons for this are not hard to find. Samoff (1999:262) suggests that cause and effect are very difficult to establish clearly in education and that the links between education and development more generally are even harder to establish. 'However daunting the challenge,' he urges, 'research on the education sector must inquire about what education assistance programmes have accomplished, at both the smaller and larger scales. Funding agencies ought to be vitally interested in the consequences of foreign aid and of their own contributions, both positive and negative. Beyond immediate programme [objectives]…what have been the impacts of foreign assistance on education, including instructional activities, agenda setting, decision-making, finance, management, community participation, and student and parental satisfaction? What have been the shorter and longer term results? Why?'

Such issues are rarely, if at all, specified in education sector studies' terms of reference. Yet they should be. For longer term results, especially, cannot be assessed in education programme evaluation, simply because such evaluation does not normally take place in the longer term. More importantly, most governments are signatories to the agreement on the global Education for All (EFA) goals[13] reached in the Dakar World Education Forum in 2000, which expanded upon the Millennium Development Goals (MDGs) for education issued in the same year. Donor agencies are now being urged to provide more funds than ever before in support of efforts to achieve the MDGs. But there have been few, if any, studies commissioned to analyse the long term impact of foreign assistance to the education sector. On the other hand, from many years ago micro-level country studies have shown again and again that household income and parental education levels determine children's school enrolment to a greater degree than education policy interventions (Clemens 2004; Pritchett 2004). In spite of all the evidence that enrolment outcomes depend on much more than education policy and expenditures, discussion of the MDGs for education centres almost exclusively on public finance of education as the prime determinant of raising enrolment and completion rates. To be sure, there are several studies which determine the costs of universal primary education as if (government, and donor) finance was the principal determinant of growth in schooling, but this goes against the large majority of empirical evidence. Indeed, Al-Samarrai (2006) has concluded from a cross-country analysis that the link between public education expenditure and educational outcomes, as measured by a range of indicators, is weak *at best* and that the achievement of the MDGs will require more than just increases in expenditure on primary education. Furthermore, the MDGs assume a rate of growth in enrolment in primary schools that is much faster than historically has been the case in both rich and poor countries (Clemens *et al*. 2004). The MDGs also tend to ignore the fact that education does not occur in a vacuum but in a broader economic environment, and that alternative means of encouraging the demand for education, such as providing cash incentives to low-income mothers (Morley and Coady 2003), need to be included in policy discussions between governments and donors.

---

[13]   The following are the global EFA goals:
- Expand and improve comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children.
- Ensure that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete free and compulsory primary education of good quality.
- Ensure that the learning needs of young people and adults are met through equitable access to appropriate learning and life skills programmes.
- Achieve a 50 percent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults.
- Eliminate gender disparities in primary and secondary education by 2005, and achieve gender equality in education by 2015, with a focus on ensuring girls' full and equal access to and achievement in basic education of good quality.
- Improve all aspects of the quality of education and ensure excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.

Project and evaluation design. The extent to which the evaluation process can be successful and beneficial to all stakeholders is dependent upon (i) good project design, (ii) substantial in-country capacity for *planned* regular monitoring and evaluation activities, and (iii) access to reliable hard data.

In many projects evaluated, project design was found to be weak. For example, in Tanzania (Mercer *et al*. 2000) the LogFrames of all three projects had been constructed by individuals and had not been developed jointly through workshops with stakeholders at central level, let alone at district or institutional levels. In addition, the consultants who devised the LogFrames did not appear to have been well-versed in the rules of logic which should be applied.

In the official donor agency guidelines on project design and evaluation, there is considerable emphasis on stakeholder participation. However, participation is not straightforward and needs to be considered carefully. Ebbutt (1998) and Riddell (1999) offer notes of caution which reflect this author's experience. Firstly, participation is cultural: at the field level, communities may participate in the planning of a project but are often uncritical of the efforts of those bearing gifts, thus allowing the chance for the ideas of planners to override those of potential beneficiaries. Secondly, there is a cultural difference in the use of language: external evaluators rely heavily on documentation, but often the reality for local communities is of an oral tradition in a language other than that of the outsider. Thirdly, insufficient attention has been given in project design and in practice to development of national and local capacity for follow-up, monitoring and evaluation of project activities (Mercer *et al*. 2001c). This leads to imbalances between the "expert" and the uninitiated and to a narrow conceptualisation of design, monitoring and evaluation activities which are more likely to be according to a pattern dictated by the donor or "agent". It must be remembered that, for the majority of stakeholders, education project design and evaluation are not part of their regular day-to-day activities. Stakeholders thus need a measure of training to avoid a skew towards the needs of policy makers rather than practitioners or beneficiaries.

For the EC, as for other donors, the first full assessment of a project is the mid-term review. This sometimes takes place towards the end of a project and leaves limited opportunity for remedial action.[14] To be sure, when the timing of the review is right, its findings and recommendations may lead to adjustment of the project under the next financing period. However, during our field visits (Mercer *et al*. 2002b) we found a number of shortcomings in the reporting of education and training projects. These are summarised in Box 1.

Box 1: Shortcomings of EC supported education project reports analysed during field visits between November 2000 and February 2001

- Input-oriented
- Focused on financial and administrative issues
- Laid out according to contractor rather than project or component
- Little indication of the degree of involvement and participation of beneficiaries
- Not approved by the major direct beneficiaries
- Description of, and justification for, deviation from the Work Plan not included
- No analysis of results achieved
- Factors influencing project results not indicated
- Lessons learned not included
- Difficult to locate in Delegation filing systems

The shortcomings of education project reports are exacerbated by a lack of institutional memory caused by the absence of handover reports when people leave a project or the Delegation. This severely limits evaluation of a project's effectiveness, and especially evaluation of its impact which can be carried out some years after the end of the project. Evaluation requirements, such as control groups, benchmark studies and systems for collecting and analysing information during the project, need to be incorporated into project design, thus allowing for an ex-post assessment of

---

[14] This situation improved with the establishment of the EC monitoring programme. Under this scheme, education support programmes/projects in all countries where the EC is active are visited and monitoring reports are available to desk officers and delegations through the EC intranet system.

project impact. The team did not find any evidence that effectiveness or impact evaluation was planned or budgeted for in any of the projects examined. This meant that the EC, as a donor supporting education and human resource development, had limited capacity to learn from experience or to carry out special studies on intervention models that were innovative or promising in terms of potential results and impact.

Although regular monitoring and evaluation activities had not been incorporated into project design in most projects reviewed in depth[15] in Mercer *et al.* (2002b), it is important to point out that a system for government-led education sector performance monitoring, illustrated in Box 2, had been successfully established in Uganda by the time of our field visit in November 2000. In effect it had been strongly encouraged by DFID and other bilateral agencies in order that both donors and government might take a sector-wide approach (SWAp) to educational development.[16]

Box 2: The education sector performance monitoring system in Uganda, 2001

➢ A medium term budget framework (MTBF) reflecting the targets set out in the Education Strategic Investment Plan for the education sector to receive 31 percent of recurrent government expenditure of which primary education is to receive 65 percent;

➢ six-monthly education sector reviews with donors and stakeholders, one of which is linked with the MTBF review;

➢ a set of (a) *critical undertakings* or targets for government's financial commitment, public expenditure management and quality enhancement—the last of which having, for example, specific targets for pupil/teacher, pupil/textbook and pupil/classroom ratios—and (b) *process undertakings* decided upon at each sector review—referring, for example, to equitable access, teacher recruitment and monitoring and evaluation;

➢ an Annual Audit carried out by the Office of the Auditor General and accepted by all donors;

➢ an education management information system (EMIS) which allows for participatory monitoring; and

➢ an ample and structured *consultation* mechanism based upon:
  • weekly Ministry of Education and Sports Top Management meetings; and
  • meetings every two months between representatives of the ministries of education, finance, local government and public service, and of donors, NGOs, civil society and education institutions.

If a donor agency is an active signatory to this sort of performance monitoring system, the evaluation of programmes or projects financed by that particular agency becomes straightforward and participatory, provided that there is an effective education management information system (EMIS) in place. Unfortunately, this is rarely the case. To be sure, in ministries of education (and in regional and district offices) there is commonly an office which is responsible for planning and/or evaluation and for the administration of a so-called EMIS. Indeed, it is likely that there is a computerised **system** in place which deals with regular **education** statistics. But the *heart* of the EMIS is weak or does not exist, i.e. there is no **management information** or it is not effectively used for decision-making at any level. The evaluation process is the worse for it.

---
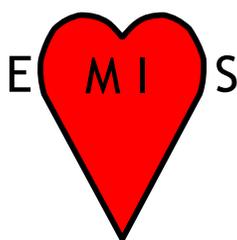
[15] In-depth analysis was carried out for 16 ACP countries—Antigua, Botswana, Burkino Faso, Dominican Republic, Ethiopia, Ghana, Guinea Conakry, Mali, Mozambique, Namibia, Papua New Guinea, Tanzania, Uganda, Vanuatu, Zambia and Zimbabwe—representing ca. 50% of total EC funding planned for the education sector in ACP countries, 1993-2000.

[16] In 2001 the Ugandan education sector satisfied all criteria for investment in a SWAp , namely: (i) a clear sectoral policy document; (ii) a coherent strategic framework; (iii) a realistic medium term sectoral expenditure framework; (iv) a comprehensive annual budget; (v) effective government-led donor coordination; (vi) jointly agreed annual targets to be met or "undertakings" to be completed; and (vii) regular joint government, donor and NGO reviews of sector performance.

The evaluation process. Experience in the final evaluation of EC support to primary education in India (Mercer *et al*. 2002a) has proved a good source from which to draw conclusions on the evaluation process itself.

The terms of reference (TOR) for the evaluation and other documented requirements displayed an excess of optimism about the extent and reliability of the hard data to which the evaluators would have access. They also needlessly corralled the definition of "evidence" within a narrow positivistic framework which equated evidence with "what can be quantified". This appeared to take no account of the substantial critical literature on the limits to quantitative data, nor of known weaknesses in India's District Primary Education Programme (DPEP) dataset, nor of the considerable advances in qualitative social research methodology, including the use of triangulation to test and validate opinion-based findings.

There were serious ambiguities about whom the evaluation was for, and overall it was expected to speak to too many audiences, some of whom had conflicting needs and expectations. Thus, the exercise was expected to meet the requirements of (i) the two states involved, Madhya Pradesh and Chhattisgarh, (iii) the Elementary Education Bureau (EEB) and the Government of India (GOI), (iii) the EC Delegation, (iv) the EC Education Programme Office (ECEPO), (v) and the EC in Brussels. Had the needs of these bodies coincided there would have been no problems, but they did not.

Whatever was claimed to the contrary, and despite the best efforts of the team itself, the evaluation found itself constantly being forced into the joint review mission (JRM) modality.[17] At district and state levels it was assumed that the evaluation was a JRM, and there was the usual JRM panoply of large set-piece meetings at which genuine interaction was frustrated either by the number of people involved or by defensiveness, and of the JRM "royal progress" round too many schools in too short a time.[18] The desire to *impress* which attended the JRM also produced an urge to *suppress* the unpalatable in terms of evidence. This certainly happened with the evaluation. The team worked hard to circumvent these problems, but they were aggravated by the overall framework of the evaluation mission, which was pure JRM: district wrap-up meetings, a state report and wrap-up, a meeting with the EEB attended by pretty well everybody, all in a high state of nervousness about preserving face, and so on.

As with the JRMs, this put the states and districts disproportionately under the spotlight, and tended to imply that if there were problems in DPEP implementation their roots lay at state level and below rather than at national level or at the interface between national and state levels. This was not only inaccurate; it was also unjust to the states and those at the cutting edge of DPEP who were struggling to implement the ambitious DPEP vision in circumstances which were always difficult and were sometimes very different (e.g. in terms of resourcing and support) to what was claimed at the higher levels. Had the evaluation not been forced into the JRM modality it might have been possible to include what had been conspicuously absent from JRMs to date, namely a balanced evaluation of the contribution to the successes and problems of DPEP of national agencies such as the EEB, the National Institute of Educational Planning and Administration and the National Council of Educational Research and Training, as well as of the states and districts themselves.

Similarly, though evaluation of EC support for primary education in India properly implied questions about the management of the EC's operations there, it was difficult to pose these when it was made clear that the EC itself and especially the Delegation (though not the ECEPO) were outside the frame of reference. The team was also subject to frequent interventions from the Delegation about what it should say, and how. Such interventions appeared to show little regard for the professional expertise of the team's members and came close to compromising the independence of their findings and judgements. Of course, it is always difficult for the body commissioning an evaluation of its own affairs to countenance with equanimity the degree of

---

[17]   DPEP began in 1995 in 42 districts spread across seven states. By November 2001, DPEP was covering 273 districts in 18 states, involving more than 51 million children and 1.1 million teachers. Throughout its implementation there were biannual joint GOI and external agency review missions. The external agencies involved in—and funds in Rs. billion, approved by the GOI Expenditure Finance Committee (EFC), committed by August 2001 to—DPEP are the World Bank/IDA (49.51), the UK DFID (9.29), the EC (5.85), the Netherlands (2.15) and UNICEF (0.36). GOI and the state governments had EFC approved commitments of Rs. 10.29 billion by August 2001.

[18]   Ebbutt (1998) also found that the external evaluator is often taken to the best sites or examples, and is treated as an honoured guest.

evaluative independence which is required for the job to be done properly. But this is a familiar enough problem in programme evaluation, and there were other, and better, ways that the relationship between the parties could have been conceived.

The evaluators were asked to write too many reports and in the wrong order. Once again this may well have been the stranglehold of JRM thinking. An immense amount of time went into the many drafts of the two state reports. In fact, gathering data for and writing these two documents dominated the entire evaluation programme, when what should have been uppermost in the minds of the team was the evaluation report for the EC, which was what the exercise was supposed to be directed towards. Any feedback to the states and the EEB should have been subsidiary to, and a by-product of, this task. Instead, it was the other way round. The state reports took most time and were endlessly redrafted and polished, the EEB aide memoire was constructed in some haste, and the final report was written under even greater pressure. Moreover, whereas the team was together during the writing of its first three reports and therefore able to interact on a constant daily basis about what these reports should say, consultation about the final report could only take place by e-mail. This is not an adequate substitute for the kinds of discussions which should have informed the production of such a document. Of course, it could be argued that the state reports provided the main documentary basis for the EEB aide memoire and EC final report, and it is true that the final report drew upon them quite heavily. However, the fact that the state reports were written not for the EC but for state and EEB functionaries meant that (a) they were politically slanted and constrained in ways that the final EC report should not have been, and (b) they contained a great deal of detail which might have been of interest to the states but was of little interest to the EC. A strategy which requires evaluators to spend such a large proportion of their time writing reports for bodies other than the one which commissioned the evaluation is of dubious wisdom.

Most of the problems above could have been avoided if the evaluation team had been consulted, well in advance, about the evaluation. The team contained considerable and acknowledged expertise not just in the substantive areas relevant to DPEP, such as planning, management, finance, primary education, pedagogy, teacher training and equity, but also in programme evaluation design and implementation. This expertise was ignored, and instead the team was presented with a flawed evaluation design which it was required to implement as it stood. It is now widely recognised that the best evaluations are those in which there is close consultation, from the outset, between those who commission an evaluation and those who undertake it. By this means the terms of reference, timetable and methodology can all be devised in a way which gives the objectives of the evaluation the best chance of success, and any ambiguities and conflicts of interest can be confronted and, ideally, circumvented. For it has to be said that matters such as those raised above—about concepts, timetables, the dynamics of state and district visits, meetings at state and national levels, evaluation design, the debates about quantitative and qualitative data, triangulation, evaluation ethics, the relationship between the control of evaluation and its content, problems of reporting and audience, and so on—were not unique to this particular evaluation but have featured in evaluation literature, especially in the USA, since the 1960s. Many of the problems of the evaluation assignment could have been avoided, or at least alleviated, had the team been brought into full partnership with the EC from the outset. And once a mutually acceptable framework had been agreed the team should have been left to implement it, rather than being subject, as it was, to constant checks that it was doing as it had been told.

There was also the problem, familiar from the EC's approach to DPEP JRMs, of last minute decisions and notifications. The team were given the draft TOR just one week before the start of the evaluation. Some of the team members were given little more notice than this of their involvement. Several of the team members had senior executive responsibilities elsewhere and all of them had a full programme of other work, including commitments entered into many months before the EC's request that they participate in the evaluation. It was unreasonable of the EC to expect them to drop everything at such short notice and make themselves available for nearly five weeks of uninterrupted fieldwork and a further extended period of preparation and writing. It was (and still is) unclear why the EC was unable to act as other donors and make its missions and evaluations the subject of proper forward planning so that those likely to be involved were able to programme their EC consultancy without jeopardising their other responsibilities. This was a

matter of efficiency as much as courtesy. After all, the EC had known since 1994 that it would need to commission an evaluation!

Finally, as with the evaluation of EC support to education in ACP countries, there was a sharp focus on financial indicators, largely internal to the EC, which served to limit the potential for local capacity building in performance monitoring. Yet, in a recent evaluation of the World Bank's current efforts in non-formal education for adults (Oxenham 2006), a major finding is that priority has been given to implementation, local ownership and local capacity building at the expense of actually developing reliable systems of monitoring, evaluation and research. As a result, educators, education policy makers and education planners still lack incontestable empirical guidance on almost all aspects of design, methods, implementation, costs and outcomes.

EVALUATION: TO WHAT EFFECT?

DFID output-to-purpose reviews are mid-term formative evaluations and as such contain comments and recommendations for action in the immediate and near future, agreed by both government officers responsible for the project and project team members. On the whole such evaluation has a positive effect.

For the most part donor agencies already have funds allocated to countries, to priority sectors and, sometimes, to specific programmes by the time an evaluation report is accepted. The need to spend—and the need to spend in a particular way: through project aid, through programme aid (e.g. general budget support) or through a SWAP where the government of a country designs an overall strategy and donors sign on to fund the sector, not individual projects—is seen to override conclusions drawn in any evaluation report. For example, following submission of the belated mid-term evaluation of support to Tanzania's ESDP in which it was pointed out among other things that the LogFrames had been poorly designed and by individual consultants alone rather than together with major stakeholders, the EC asked the evaluator to design by himself yet another LogFrame for expenditure of the remaining EC and UK funds. (This was not quite the effect that the evaluator expected his report to have! He declined to carry out the task and recommended that the remaining funds be pooled with other funds allocated to the country for use in better planned programmes.)

Generally speaking, donor agencies give little, if any, feedback to contracted external evaluators after their final report has been accepted. It is rare that such evaluators learn what effects their evaluation has had on subsequent aid to the education sector, especially if they cease to be involved with the funding agency in the country where the evaluation has taken place. Programme evaluation is considered to be something that ought to be done as a matter of course and ensuing programmes will have been pre-determined to a large extent before evaluation of existing programmes takes place. This was the case with the evaluation of EC support to primary education in India. The evaluators recommended that the proportion of funds to be managed directly by the EC (rather than by GOI or the state governments) in the proposed programme of support to basic education—primary and junior secondary education—be re-examined in light of the less than satisfactory efficiency with which funds had been managed directly by the EC in the programme being evaluated, where GOI and the state governments were considered to have managed their proportion of funds efficiently. Since this proportion (5%) had already been decided by the EC, the team's recommendation had no effect.

In reality, the effort that an agency expends on evaluating its activities in the education sector is useful only to the extent that the findings are used to inform future work. The task is to distil lessons from past practice that offer relevant guidance for future action. The challenge is not necessarily to do more evaluation, but to ensure that the evaluations which are done are strategic, planned from inception and conducted at an appropriate time and in an appropriate manner, and that the findings are shared in effective ways.

CONCLUSION

The lessons learned from experience in the external evaluation of education programmes are many but can be summarised as follows.

- The design of what needs to be monitored and evaluated and how evaluation can best take place should be an integral part of project or programme design, and not an afterthought. The evaluation process also needs to be fully costed for all stages of a project, including, ideally, ex-post evaluation in which the longer term impact can be assessed. An important corollary of this is the requirement for continued regular data collection and analysis throughout a project and becoming a matter of course beyond the project.

- Essential for both formative and summative evaluation are baseline measures of *valued* results and outcomes emanating from a project. Studies which establish these baseline measures also need to be designed, costed and paid for.

- Stakeholder involvement, at local, regional and national levels, in both programme and monitoring and evaluation design is essential if local realities and needs as well as visions are to be linked closely with national reform.

- The focus of evaluation needs to go beyond quantitative measures, especially narrowly financial measures. By putting a price tag on the outcomes of an education programme, cost estimates inadvertently create an illusion that any objective can be met, if only the right amount of money can be mobilised. Quantitative survey techniques can best be used where there is a need to draw general conclusions, whereas qualitative studies are required in order to understand particular cases in depth.

- Developing local capacity in monitoring and evaluation is important for broadening accountability to others beyond specific funding agencies, to stakeholders and direct beneficiaries. Too often in education programmes the *raison d'être* for ministry of education staff—the learners—is forgotten. Equally important, of course, is the need to establish and develop reliable and sustainable systems of monitoring, evaluation and research through which the evaluation process can be treated as a learning experience equally for evaluators and others involved.

The question 'Evaluation: to what effect?' remains largely unanswered. To do so in any meaningful way requires considerable research by aid agencies in the form of studies to determine the relationship between programme evaluation and future policy and activities in the education sector. Equally important is the need for systems to be in place where evaluation reports can be shared between governments, donor agencies and concerned stakeholders and where, as a result, better informed decisions can be made to ensure greater participation in education and higher quality and more relevant learning outcomes.

REFERENCES

ADB. 2003. *Program Completion Report* (PCR: MON 27371) on the Education Sector Development Program (Loan 1507/1508-MON[SF]) in Mongolia. Asian Development Bank: Manila.

ADB. 2005. *Guidelines for the Preparation of Country Assistance Program Evaluation Reports*. Draft: December 2005. Asian Development Bank: Manila.

Al-Samarrai S. 2006. Achieving education for all: how much does money matter? *Journal of International Development* **18**(2).

Clemens M. 2004. The long walk to school: international education goals in historical perspective. *Working Paper* 37, Center for Global Development: Washington, DC.

Clemens M, Kenny C, Moss T. 2004. The trouble with MDGs: confronting expectations of aid and development success. *Working Paper* 40, Center for Global Development: Washington, DC.

Ebbutt D. 1998. Evaluation of projects in the developing world: some cultural and methodological issues. *International Journal of Educational Development* **18**(5).

European Commission. 2001. *Evaluation in the European Commission: a guide to the evaluation procedures and structures currently operational in the Commission's external cooperation programmes*. European

Commission EuropeAid Cooperation Office: Brussels.

Mercer M, White K, Katunzi N. 2000. Joint mid-term evaluation of EC, DFID and Ireland Aid support to the Education Sector Development Programme design, preparation and management process. *Final Report*. European Commission: Dar es Salaam.

Mercer M, Craig R, Morrison M, Jallow M. 2001a. The Gambia: Education Management Project. *Output to Purpose Review*. Department for International Development: London.

Mercer M, Gosparini P, Orivel F, Kayonga S. 2001b. Evaluation of EC Support to the Education Sector in ACP Countries. *Uganda Country Report*. European Commission: Brussels.

Mercer M, Melchiori P, Sirtori M, Sikoska T. 2001c. Evaluation of EC Support to the Education Sector in ACP Countries. *Dominican Republic Country Report*. European Commission: Brussels.

Mercer M, Alexander R, Ramachandran H, Rao P, Singh A.K. 2002a. Final Evaluation (Part I) of EC Support to Primary Education in India. *Final Report*. European Commission: Brussels.

Mercer M, Stenbäck T, Orivel F, Melchiori P, Gosparini P, Sirtori M. 2002b. Evaluation of EC Support to the Education Sector in ACP Countries. *Synthesis Report*. European Commission: Brussels.

Morley S, Coady D. 2003. *From Social Assistance to Social Development: Targeted Education Subsidies in Developing Countries*. Center for Global Development and International Food Policy Research Institute: Washington, DC.

Oxenham J. 2006. *Education for All Adults: the Forgotten Challenge*. World Bank Directions in Development Series. World Bank: Washington, DC.

Passingham S, Mercer M, Little A, Boyle, S. 1998. Sri Lanka: Primary Education Planning Project. *Output to Purpose Review*. DFID South East Asia: Bangkok.

Pritchett L. 2004. Towards a new consensus for addressing the global challenge of the lack of education. *Working Paper* 43, Center for Global Development: Washington, DC.

Riddell A. 1999. Evaluations of educational reform programmes in developing countries: whose life is it anyway? *International Journal of Educational Development* **19**(6).

Samoff J. 1999. Education sector analysis in Africa: limited national control and even less national ownership. *International Journal of Educational Development* **19**(4-5).