

Macro-level evaluations and the choice of aid modalities

Tony Killick¹

Orientation

In recent years, the pendulum of professional opinion about effective aid modalities has swung away from an original concentration on project-based assistance in favour of more programmatic forms, most notably budget support and the associated modality of debt relief. Although some donors remain wedded to the project mode, there is little doubt about the direction of trend. Thus, a recent British policy statement on the subject (DfID, 2000:93):

...there needs to be a real improvement in the way that assistance is delivered. That means reducing support for stand-alone projects, and increasing support for sector-wide reforms. Where governments have a strong commitment to poverty reduction and strong policies in place, it means moving towards providing financial support directly to recipient government budgets using their own systems.

Since the end of the 1970s the World Bank has, of course, been providing programmatic 'policy-based' adjustment credits but, in low-income countries, the more recent opening of its Poverty Reduction Support Credit (PRSC) window has added a further dimension, not least because -- to its great credit -- the Bank is beginning to go to some lengths to harmonise the terms of its PRSCs with those of budget support baskets put together by bilateral and other donors. The introduction of the Enhanced HIPC debt relief scheme in 1999 added a further string to the programme aid bow.

It is hardly contentious to suggest that this shift represents a challenge to evaluators. New issues arise and new needs for evaluation are created. One 'big' issue is the extent to which the shift from project to programme support is soundly based on evidence about the comparative effectiveness of different modalities: have past evaluations and research provided the evidence needed for informed decisions about modalities? Another issue, however, is the extent to which those who make decisions about modalities are actually driven by evidential factors: do they listen to evaluators? Perhaps to some extent they are unable to because of time-lags. The contemporary aid environment is a rather fast-changing one whereas evaluation is retrospective, occasional and reflects the aid choices of the past rather than the fascinations of today. Finally, the switch to programmatic support raises questions about the level at which

¹ Senior Research Associate, Overseas Development Institute, London and Visiting Professor, University of Surrey. I am grateful for helpful comments on an earlier draft by Howard White and others at OED.

evaluation should be conducted and who should undertake it.

In the following, I want to illustrate some of these issues by reference to four specific topics, all broadly macro in the sense of 'big'.

First and starting from the common assertion that programme aid is superior to the project-based alternative because it lowers **transactions costs**, I want to raise questions about the state of our knowledge about comparative costs, and about the loose way in which this concept is called in aid in the rhetoric of debates about aid modalities.

Second, I want to utilise a recent OED study to argue that, in aid-effectiveness terms, there is probably a rather severe conflict between evidence-based considerations and the heavy weight attached to political factors in the use of aid resources for the provision of **debt relief**. In this case, the evidence is there ... but the question is, Who's listening?

My *third* illustration also relates to an apparent -- but this time much more complex -- failure of decision-makers to take an evidence-based approach, this time not only because of 'politics' but also because of institutional factors and because of research time-lags. This refers to interactions between work on **policy conditionality** and the practices of the BWIs.

The logic of the movement towards partnership-based general budget support is that more and more evaluative work needs to be done at the country-wide, as against project- or perhaps sector-specific, level. This not only has large implications for the types of skill and information required, and for methodologies, but it also raises the less obvious question, who should undertake such studies? My *fourth* section, therefore, will draw on a recent exercise in Tanzania to argue the case that such work will often best be conducted by evaluators independent of both donors and recipient governments -- so-called **independent monitoring groups**.

I - 'Transactions costs' and the choice of aid modalities

Programme aid is claimed by its growing number of advocates to be superior on a number of grounds (stronger influence on the policy environment, superior ownership properties, greater aggregate coherence, etc) but many of these boil down to the oft-heard claim that a given sum of programme aid gives rise to smaller transactions costs than an equivalent amount given in the form of traditional development projects. Thus, a recent report on aid relations in Tanzania (IMG, 2003:37):

...we see programme aid as avoiding various of the drawbacks ... of project aid, especially when it takes the form of a pooling of resources and co-ordination among donors. The presumption is that, with a much smaller number of reporting points, with no necessity for a multitude of Project Implementation Units, with donors accepting common reporting procedures and standards, with tied procurement virtually eliminated

and with much less necessity for enforced project-linked technical assistance activities, transactions costs per dollar of aid received will be much smaller.

That is now the conventional wisdom and I agree with it *as a hypothesis*. There are many complaints levelled against the project approach, which is seen as undermining local capacity development, imposing heavy procurement and reporting requirements because of the large numbers of individual projects, and reducing the value of aid through the imposition of procurement tying. However, the superiority of programme aid in this respect remains only an hypothesis because it rests on a presumption of the comparative costs of the respective aid modalities which has not, to my knowledge, been undertaken in any systematic way. In fact, a literature search revealed only one study, examining aid transactions costs in Vietnam (Brown *et al*, 2000).² I want to argue that the language of transactions costs has been used too loosely and would merit a substantial empirical study.

Conceptual issues

One reason for taking a careful look is that there are some conceptual issues that are generally disregarded. The text-book meaning of transactions costs is clear enough:

the costs of negotiating, monitoring and consummating a contractual arrangement

However, there is a lack of clarity in professional usage about what the chief categories are in the case of aid-delivery contracts. There are questions too about the *distribution* of such costs, as between donors and recipients, within the structures of each party, and over time. First, categories. The following is an attempted classification of types of transactions costs:

Administrative (recipient and donor)

- *Ex ante* identification, appraisal and preparation of the aided activity
- Implementation, including any special procurement requirements
- Monitoring, administrative and financial reporting, including special arrangements to safeguard against maladministration and to secure aid effectiveness
- *Ex post* evaluation, including tracker, outcome and impact studies

(To be included in each of the above items are the administrative opportunity costs resulting from absorption of scarce staff time within both donor and recipient agencies, although the presumption is that the burden of such costs is

² In addition, see Annex 3 of an OED report on aid co-ordination (1999) which contains a brief discussion of the costs of co-ordination.

disproportionately high within recipient administrations).

As a special case, there are the opportunity and other costs arising from 'staff capture' as a result of donor inducements to recipient civil servants to give priority to a particular aided activity, or to move across to that activity from regular duties.

Tying (recipient)

- Costs arising for recipients from loss of coherence and national ownership as a result of the tying of aid to large numbers of individual projects
- The higher costs of imports resulting from procurement tying. The World Bank (1998) has estimated that this typically reduces the value of aid by about a quarter.
- Costs arising for recipients from losses of national ownership and sovereignty as a result of policy tying (conditionality), and perceived increases in political risks resulting from any associated governmental loss of policy freedom (but see Section I on conditionality).

Fiscal (recipient)

- Erosion of fiscal discipline by extent of off-budget allocations (associated with project approaches) and the wider consequences of that.
- Fiscal management problems generated by under-funding, requirements for counterpart funding and the recurrent costs of aided activities.
- Planning problems created by the volatility and unpredictability of different classes of aid, and the macroeconomic consequences of that.³

I shall suggest shortly that programme aid is apt to give rise to more transactions costs than is commonly assumed but what is already clear from this listing is that all aid modalities are likely to give rise to a number of these different cost types. However, they do not do so equally and it is easy to see from this categorisation why project aid has a bad reputation.

Before looking at the programme aid case, we should consider how different forms of delivery distribute transactions costs differently among the various parties. Take the case of a shift from project to macro-level programme aid. Within the donor agency in question, there will be a shift in administrative and professional work away from project specialists in favour of more macroeconomically oriented staff (which is one reason why project staff so often oppose the move to programme aid and why this can be a sensitive issue in the internal politics of donor agencies). Since a lot of donor project-related administrative costs are often shifted to specially-contracted consultants, whereas this

³ Conceptually, this might be thought of as the opportunity cost of the extra volume of reserves it would be necessary to hold in order to smooth out unexpected variations in aid receipts.

may be less readily done with programme aid, a shift in favour of programme aid is apt to increase the work burdens of aid agency staff, as well as concentrating more of it in agencies' 'macro' departments.

Shifts will also occur within the recipient administration. In crude terms, more resources -- and the associated transactions costs -- flow through the budget from 'the centre', "the Ministry of Finance and Development". It is this ministry which will have to bear the main brunt of donors' reporting and evaluation requirements, so that its capacity to meet these becomes a critical variable. Even line ministries may also find themselves with additional burdens, because they are now having to account for items, through regular budgetary processes, which previously were off-budget and most likely being handled semi-autonomously by Project Implementation Units and the like.

Another consideration is the way comparative transactions costs may change over time. It is suggested, for example, that some of the transactions costs associated with programme aid are essentially start-up costs, which can be expected to diminish as experience is gained and systems are developed, and that while a shift to programme aid may well eventually reduce aggregate transactions costs, the initial effect may be to raise them (Lawson *et al*, 2002).

A final complication concerns the fact that at least some of the costs described above are intended to produce benefits. This is most obviously the case with the various administrative measures intended to safeguard against mis-use of aid. Similarly, policy conditionality is intended to improve the policy environment and hence aid effectiveness. On the one hand, it obviously does not make sense to view the costs in abstraction from whatever benefits they might generate. On the other hand, since we are not here talking of costs and benefits which can be reduced to a common quantifiable numeraire, a large element of judgement is involved in any attempt to net them out.

It is considerations like this which help to explain why researchers and evaluators have in the past shied away from the direct study of transactions costs. As the Brown *et al* (2000) paper spells out, there are numerous other difficulties to overcome. Various of the costs listed earlier cannot be directly observed. Even some of those which can be observed -- for example, some of the fiscal costs -- are difficult to do more than describe. The totality of aid transactions in aid-dependent economies involves large numbers of agencies and individuals, with the result that the information required to get a view of the extent of transactions costs will be fragmented and dispersed. There will also be a variety of perceptions. One man's cost is another's safeguard.

All this is daunting. Evidently, no more than indicative evidence would be possible from even the most thorough investigation. At the same time, the concept of transactions costs is being used to justify really large changes in the modalities of aid delivery, which suggests that it would be well worthwhile to strip away as much as possible of the conjecture which surrounds this topic. At present, our ignorance is such

that a statement like “programme (or project) aid is preferable because it reduces transactions costs” is not much better than saying “I like programme (or project) aid.” In particular, the empirical study of transactions costs seems worth-while because there are grounds for questioning the alleged superiority of programme aid in this regard.

Costs of programme aid

First, we should note a major type of cost which is largely unique to programme aid: what is described above as policy-tying, or conditionality. As indicated there, conditionality, *assuming it achieves its intended purpose* (but see Section I above), can be thought of as generating costs because, by requiring governments to undertake policy measures they would not otherwise have chosen – or to do so at a faster pace – they are apt to generate political risks and to reduce governments’ ability to manage these risks. Note the risks are born almost entirely by the recipient, as distinct from the BWI (or bilateral donor) in question. There is also likely to be a perceived loss of sovereignty and, therefore, of national ownership of policies. Many policy changes also generate adjustment costs of various kinds.

It may reasonably be objected that in contemporary best-practice programme aid is seen as going along with a ‘partnership’ approach, which has precisely been designed as an alternative to extensive policy conditionality. However, whatever may be the case on the bilateral front, we have seen in Section I that the BWIs have thus far not introduced any substantial reduction in reliance on conditionality. Even among bilateral donors, it is clear that far from all of them have ‘bought’ the partnership approach as an alternative to conditionality (e.g. see Lawson *et al*, 2002). Although there are some that are serious about moving in that direction, it is probably true to say that most continue to have faith in conditionality, especially the use of prior actions. Of course, it could also be objected that conditionality is for the most part ineffective and, therefore, that most of the costs in question do not actually occur. This may well be the case but it is scarcely evidence of the superiority of programme aid that one of the principal ways in which it is expected to be made effective does not actually work very well.

Conditionality, of course, necessitates monitoring and reporting, which brings us to the administrative transactions costs listed earlier. Of course, project aid also generates administrative costs but it is by no means clear that these are proportionately larger than with programme aid. This is well illustrated by the elaborate monitoring arrangements currently being established in connection with the HIPC initiative. Here the essence of the task is to ensure that extra monies released through debt relief are spent on the designated priority poverty-reducing budgets but to do so is actually a complex task, involving significant costs. Some bilateral donors also get into the, surely ill-conceived, business of trying to track what ‘their’ funds are spent on through the budget, imposing substantial transactions burdens both on themselves and recipient administrations. As White and Dijkstra (2003:476-77) put it:

Donor attempts to trace their funds resulted in complicated systems

which were anyhow likely to be an accounting fiction. These attempts also led to systems trying to micro-manage the recipient's foreign exchange allocation system which may have been inappropriate ... Some agencies may have been forced to adopt such procedures for domestic political reasons, though a better response would be to educate decision-makers on how programme aid works.

More generally, the enhanced levels of 'dialoguing' and conditionality associated with programme aid generate substantial costs for all parties, but especially for recipient administrations. Especially where there is a combination of sectoral programme aid (SWAps) and direct budget support, the tendency for the development of rather elaborate mechanisms for dialoguing, monitoring, reporting and *ex post* evaluation is apt to impose substantial burdens. This is particularly so where there is a less than perfect convergence of government and donor objectives, and less than complete trust by the latter of the former.

A further aspect of programme aid is the special extent to which it relies on co-ordination and harmonisation in order to reduce transactions costs – and the well-known difficulties of achieving these in practice. This is a sufficiently well-understood point to need no elaboration here. Since many of the costs of co-ordination and harmonisation fall on the local representatives of donor agencies, there may be a tendency to discount these as less onerous, although they are nonetheless real. However, I would argue that effective co-ordination is most likely to be achieved where the government is actively involved and 'in the driving seat', in which case quite a lot of the costs in question fall upon it. Of course, in the (more general) case where co-ordination and harmonisation are not very effective, most of the costs of these failures fall upon the recipient country.

A final reason why we should not take it as axiomatic that programme aid is superior in transactions costs terms concerns the problems it creates for fiscal management. Part of the problem here is that budget support tends to be geared to the fiscal cycles of the various donors rather than of the recipient, and gives rise to patterns of commitments and disbursements which do not fit with recipients cycles of budget preparation and execution. Even more serious, programme aid appears to be especially prone to volatility, political 'interference' and unpredictability. This is well demonstrated by a research paper by IMF staff members (Buliř and Hamann, 2001) which investigates the volatility and predictability of aid in general and of programme versus project aid. Among their remarkable results are the following:

- Aid is more volatile than domestic fiscal revenues and tends to be pro-cyclical.
- Fiscal planners are highly uncertain of aid receipts and "the information content of commitments made by donors is either very small or statistically insignificant."

- There are much larger prediction errors in programme aid than with project assistance and a stronger tendency to over-estimation. This is partly because of the application of conditionality but there are other reasons too.

(The paper just cited, incidentally, is an excellent example of how at least some aspects of transactions costs are amenable to rigorous empirical testing.)

The hypothesis that transactions costs are proportionately lower with programme aid is thus not one whose validity should be taken for granted. Indeed, one study of aid to the health sector has suggested that the truth may lie the other way round (Foster *at al*, 2000), a judgement with which Lawson *et al* (2002) concur, at least as it relates to the shorter-term.

Conclusion

Table 1 (overleaf) is an impressionistic attempt to pull together various of the suggestions made in the text. The identified cost types are those suggested earlier and the suggested extent of costs in each entry should be read as *relative to the investment cost of the aided activity in question*.

One obvious conclusion to emerge from this is that the greater part of the burden falls upon recipients rather than donors, a factor which would be multiplied manifold if it was considered relative to the resources available respectively to donor and recipient administrations. Concentrating on the recipient side of the table, it does seem to emerge that lower transactions costs may be associated with programme aid but one of the limitations of the exercise is that it treats all types of cost as of equal significance, whereas the seriousness of loss of freedom for political management or of macro destabilisation is evidently on a different scale from, say, evaluation costs. That apart, the entries under programme aid challenge any idea that this modalities generates only small transactions costs.

The main point, however, is that the entries in the table are only my guesses, which I dignify by calling them hypotheses. We really don't know.

My comments above are by no means intended as an attack on programme aid. For one thing, there are other reasons for preferring it over project approaches. But what I have tried to do is to suggest that the trend in aid share, away from projects in favour of programme aid may not be as strongly based on evidence as is commonly assumed and that, to some extent, the language of transactions costs has been used as a weapon in the rhetoric of debates about the effectiveness of alternative modalities. There is, therefore, a strong case for a serious evaluative study of the whole transactions costs question. Because it would have to be a major effort, such a study would probably lie outside the capabilities and budgets of the evaluation departments of bilateral donors, which leaves

the World Bank's OED as the best equipped, perhaps heading a consortium of donor evaluators.

Table 1 -- Hypotheses about comparative transactions costs of programme and project aid

<i>TRANSACTION COSTS</i>	<i>RECIPIENT</i>		<i>DONOR</i>	
	<i>programme</i>	<i>project</i>	<i>programme</i>	<i>project</i>
<i>Administrative</i>				
Pre-investment appraisal etc	Moderate	Moderate	Moderate	Very large
Implementation	Low	Potentially large	Low	Moderate
Monitoring & reporting	Large	Very large	Large	Moderate
Ex-post evaluation	Low	Moderate	Large	Moderate
<i>Tying</i>				
Loss of coherence & ownership	Nil (?)	Large		
Price premium	Nil	Large ("25%")		
Political etc costs of policy conditionality	Potentially large	Low		
<i>Fiscal</i>				
Counterpart fund & recurrent cost problems	Nil	Large		
Volatility/unpredictability, destabilisation	Very large	Moderate/large	Moderate	Moderate
Erosion of budget by extra-budgetary aid	Nil	Large		

II – The OED’s HIPC evaluation: who’s listening?

My second topic raises questions -- to be taken further in a later session of this conference -- about the capacity of evaluators to induce changes in policies and projects where decision-makers appear to place a low value of evidence-based considerations, in this case the provision of debt relief for low-income countries. The OED has recently completed an evaluation of the HIPC debt initiative (OED 2003). This examines issues of design, implementation and (in a necessarily preliminary way) effectiveness. While it does not overtly spell out the implications of its findings for the likely developmental effectiveness of the assistance in question, the implications are barely below the surface and are negative. It can be inferred from the report that the HIPC scheme, while there are important items to put on the credit side of its ledger, has been bad news for aid effectiveness, for reasons summarised below, although I will take the argument further than the OED report chooses to.

Failure of the additionality principle

The OED report valuably draws attention to the central importance, when assessing the effects of the Enhanced Heavily Indebted Poor Country debt initiative (hereafter E-HIPC, sometimes known as HIPC-II), of the notion of *additionality*, i.e. the extent to which resources provided by creditors to finance the E-HIPC scheme are additional to normal and previously planned flows of development assistance. The creditors formally agreed the principle that they should be additional and, had it been observed, the result would have been a substantial net addition to the total volume of resource transfers to poor countries. However, the OED report draws attention to the facts (a) that the E-HIPC scheme contains no specific provisions to ensure that additionality is observed and (b) that the early evidence does not suggest that it has been achieved in practice. Indeed, introduction of the original HIPC scheme in 1995 coincided with a very sharp fall in total transfers to HIPC countries, to levels which have not since recovered.

Although additionality is an elusive concept, because of the difficulties of assessing the counter-factual, what can be said is that successive HIPC schemes have not prevented a decline in net financial transfers to debtor countries, because the financial cost of debt relief has apparently been deducted from intended aid transfers (see also Gunter, 2001, for a similar finding), and that actual transfers to them have been well below projected levels. It would take a major creditor-by-creditor research effort to be able to estimate the true extent of additionality but the presumption is that, in the general case and over the longer-term, a creditor-donor government will decide how much in total it wishes to provide as assistance and will then reduce its other aid programmes to accommodate the budgetary cost of debt relief.

Inefficient redistribution of aid

Despite the stagnation of aid to the HIPC countries, there has nonetheless been a sharp redistribution of resources away from low-income countries which are not rated as eligible for HIPC relief. They have been cut particularly heavily. Comparing 2000

(latest available) with 1998 indicates the following changes in the respective shares of these two groups (% of total net resource transfers):

	1998	2000
<i>HIPC</i>	44	76
<i>non-HIPC</i>	56	24

Evidently, recipients of E-HIPC assistance are being treated in a privileged way, but is this consistent with aid effectiveness? The E-HIPC scheme has the reduction of poverty as one of its principal objectives but eligibility for HIPC relief is not well correlated with the incidence of poverty. Outside Africa, China, India, Bangladesh, Pakistan, and Indonesia are obvious examples of non-HIPC countries with much poverty. Within Africa, Nigeria stands out as among the excluded, along with Eritrea, Namibia, South Africa, and Zimbabwe -- all non-HIPC countries with large numbers living in absolute poverty. Ranis and Stewart (2001) estimate that, among low-income countries, actual and potential HIPC agreements cover only about a quarter of all people thought to be living in poverty. The distributional implications of an E-HIPC scheme without additionality are such as to raise questions about the consistency of the resulting geographical pattern of resource transfers with a poverty-reduction objective. There is a real sense in which the poor in countries outside the HIPC scheme are subsidising those who live in HIPC countries.

There are other reasons for questioning the desirability of this redistribution among recipient countries, given the failure of additionality. One flows from the erection of indebtedness as the prime criterion of eligibility for E-HIPC assistance. Hitherto, a well-established conclusion of the aid-effectiveness literature has been that aid should be allocated selectively on the basis of the quality (however defined) of prospective recipients' economic and social policies. Inserting indebtedness as an overriding criterion and giving E-HIPC debt relief a privileged status among aid modalities seriously impedes application of this selectivity criterion. Although exogenous factors have been important too, the case studies conducted for the OED report showed that poor past fiscal policies and failures to adjust were important reasons for the emergence of unsustainable debt burdens. This confirmed earlier results by Brooks *et al* (1998), and others before them, showing weak macro and adjustment policies and lack of prudent debt management strategies to be associated with debt difficulties among HIPCs. Indebtedness tends to be *inversely* correlated with the quality of past policies. Thus, the global redistribution of aid may be consistent neither with maximum poverty reduction nor productive use of the assistance. It may also raise questions of moral hazard, punishing those with good records of economic management in order to rescue less responsible or competent governments.

Such concerns have been compounded, as the OED report makes clear, by the processes which brought the E-HIPC scheme into existence. As is well known, this was a political

response by creditor governments under varying degrees of pressure from civil society organisations to be generous at the end of the millennium, as a result of successful mobilisation by groups campaigning under the Jubilee 2000 banner. One consequence was that creditor governments not merely agreed to E-HIPC, introducing considerably more liberal terms than the original HIPC scheme, but also committed themselves to a target of getting at least 20 countries within the scheme (at 'Decision Point') by the end of 2000. This target was achieved (actually 22 countries) but only as a result of what is now labelled 'the millennium rush', with 17 debtor countries being accepted as having reached their 'decision point' between July and December 2000. Inevitably, this could only be achieved by lowering required standards of policy performance. Thus, the OED report records for this group worse policy track records, weaker development programmes and poorer past records of using aid productively. Unsurprisingly, a substantial proportion of these 'rush' countries have subsequently failed to deliver on policy promises. The pace of new inclusions has since slowed considerably but the report indicates that these most recent entrants into the scheme have particularly poor policy records. It appears that E-HIPC is turning out to be a mechanism for reallocating resources in favour of countries which, taken as a group (and with important exceptions), cannot be expected to make the best use of the scarce resources on offer.⁴

Undesirable policy biases

One of the most striking aspects of the E-HIPC scheme, as the OED report points out, is its association, under pressure from NGOs, with a particular and narrow approach to the task of reducing poverty, namely the expansion of spending on social services, to the neglect of wider growth and developmental priorities. In fairness, this trend in the allocation of aid monies had already been well under way for some time but the E-HIPC initiative took this bias to another level. The report points out that, for the 13 countries for which data were available, 65% of all resources released by E-HIPC debt relief were to be devoted to social services, with 7% on infrastructure, 4% on governance and just 1% on structural reforms. Among other things, this has been associated with a sharp rise in the share of total aid in these countries devoted to the social sectors *with an almost corresponding decline in the share of aid for production services*. Moreover, HIPC progress reports indicate that *over half* of government revenues will be earmarked for social spending in future years.

This concentration on social services raises the question, to what extent will it act on the causes of poverty? Inadequate access to education and health is certainly a powerful influence. But poverty has many other causes too, notably the effects of past economic stagnation or decline, inadequate access of the poor to various forms of capital, large and growing inequalities, high demographic dependency rates, gender biases, and various forms of disempowerment and of state failures.⁵ At least as it relates to Africa, I

⁴ For a further argument along these lines see Easterly, 2001.

⁵ For a survey of the causes of poverty in Africa see White, Killick et al, 2001, Part II.

would nominate poverty-inducing social structures and various government failures as the principal primary causal factors but even if the reader disagrees with that judgement, it is harder to dispute the proposition that a multitude of factors contributes to poverty, only a few of which have much to do with any neglect of social spending induced by the necessity to service external debts. Not the least of the negatives of the E-HIPC linkage between debt relief and social service provision is that it risks diverting attention from more fundamental causes. It also pays little heed to what is universally agreed: that far more than money is needed in order to ensure that social spending will actually raise the standards of the poor, with the report drawing attention to the pervasiveness of low efficiency, poor service quality, capacity shortfalls and low utilisation within poor countries' social services. In fact, cross-country evidence suggests that expenditure levels have little influence on educational outcomes, although comparable tests for health provide more mixed results.⁶ There is also the problem of skewed access to these services. African evidence suggests that these services are overwhelmingly enjoyed by the relatively more affluent members of the population, with the poorest quintiles receiving much less than a proportionate share (Castro-Leal *et al*, 1999).

As already hinted, one consequence of the concentration on social spending is that it biases attention away from the fundamental necessity to raise economic growth and remedy structural weaknesses. It is not sufficiently realised on what soft terms the HIPC countries have obtained capital in the past.⁷ In 1999, for example, the average terms on which the HIPCs borrowed were at 3½% interest, with an average maturity of over 20 years and a 43% grant element. Moreover, in the same year, they received grants equal to more than half of their total borrowings, so that the true 'average terms' were actually much softer than those just cited. The equivalent figures for sub-Saharan African countries were even more favourable. The obvious question to ask is, how could countries receiving capital on these terms possibly run into debt-servicing problems? There is no simple answer to that question and the explanation will doubtless vary from country to country. But in the general case it is to be found in a combination of exogenous shocks; poor past macroeconomic management, worsening the already large problem of capital flight; weak domestic saving, fiscal and export performances; and low returns to past investments. In turn, low returns can be seen as reflecting often low-quality public investment decisions; deteriorating terms of trade, obstacles to market access and inflexible export structures; past and continuing policy failings, as already discussed; and low enforceability of property rights.

Excessive indebtedness can exacerbate these disincentives -- the 'overhang' effect. In principle, the negative force of the policy biases could be mitigated if the E-HIPC

⁶ The evidence is briefly surveyed in ODI, 2003.

⁷ For an elaboration of the following argument see Killick, 2003 forthcoming. I am at this point going beyond the coverage of the OED report. See also Bird and Milne (forthcoming). The source of the following debt statistics is World Bank, 2001:260-61.

scheme were to reduce the overhang problem, in which case investment and growth could be boosted. There is little controversy that in the circumstances of middle-income debtor countries, such as the Latin American debtors who dominated the debt crisis of the 1980s, the overhang could and did have such an effect. The applicability of this argument to the small, low-income countries which characterise most HIPC and where there are so many other deterrents to investment, is less clear. The empirical evidence does not show a strong connection between debt stocks and investment levels, and the direction of causality is ambiguous. The OED report suggests that an overhang effect is speculative and has not been demonstrated convincingly.⁸ It would be unwise to rely heavily on a strong positive overhang effect from E-HIPC relief.

Creditworthiness and debt sustainability, as the OED report makes clear, will best be achieved by measures which strengthen the domestic economy, and its underlying institutional bases, and which address the above causal factors. Debt difficulties are better seen as a symptom of economic weaknesses than as a cause of them. The current donor preoccupation with social spending diverts attention away from this priority and thus carries the risk that the resources devoted to this will not be effectively deployed, meaning that the goal of debt sustainability may prove unattainable. Of course, the BWIs would deny this, asserting that PRSPs should tackle all these weaknesses as part of the anti-poverty effort. But not everything can be done at the same time; attention bias is real. Much stress is now on raising the quantity and quality of social service provisions and this is liable to further increase total consumption relative to saving, when saving and investment are already too low.

Multiple and conflicting objectives

The OED report points out that the E-HIPC scheme has been burdened with multiple and potentially conflicting objectives, namely (1) debt sustainability, (2) the acceleration of long-term growth and (3) the reduction of poverty. Were enough resources to be made available, and in grant form, it might be possible to reconcile these (although the theory of policy suggests that pursuing three objectives with one instrument is unlikely to be an efficient way of proceeding) but the failure of additionality means that this condition is not satisfied. Even if it were, there are also problems of design and attention bias, as already shown.

One further distortion that has been introduced in response to the tensions between these objectives and the inadequate size of the resource envelope is a strong tendency to base the debt sustainability analyses which underpin the E-HIPC packages on highly

⁸ For a recent examination of the debt-growth connection see Pattillo *et al*, 2002. For a large sample of developing countries this finds that high levels of debt are associated with lower economic growth, operating through lower factor productivities rather than through investment effects. However, they do not investigate HIPC, or low-income, countries as a sub-sample and are notably cautious in interpreting their results for that group.

optimistic export projections. For the 24 HIPC countries analysed by OED, the projected export growth rate was more than twice the historical average for 1990-2000, almost six times the average for 1980-2000, and almost twice as fast as the growth actually achieved in 2000-01. Not surprisingly, BWI staffs have not been transparent about how these and other key macroeconomic forecasts were derived. The point here is simple: efficiency is bound to be a casualty if an aid scheme is based on systematically unrealistic assumptions.

Conclusion

The OED evaluation has thus drawn attention to some serious problems with current debt relief arrangements. Especially if non-additionality continues to dominate, the consistency of the E-HIPC scheme with aid effectiveness is seriously in question. In qualification, it should be acknowledged that there are some compensating gains to put on the other side of the ledger. As Birdsall and Williamson (2002) have argued, aid in the form of debt relief reduces the enormous burden of never-ending debt renegotiations on the usually small number of individuals who can deal with such matters in the public administrations of low-income countries, releasing them for more productive policy analysis. In effect, E-HIPC substitutes untied programme aid for tied donor projects and this too should have beneficial effects on both transactions costs and local ownership (but see Section II above on transactions costs). As the OED report states, E-HIPC also embodies best-practice in strengthening institutions and incentives for aid co-ordination, in seeking in the design of PRSP processes to strengthen local ownership, and in its greater concern for the social content of agreed policy programmes. Moreover, the PRSPs, around which E-HIPC relief is mobilised, have pushed poverty up recipient governments' policy agendas and also, through their emphasis on participatory approaches, has the potential for involving citizens and civil society much more in policy formation, monitoring and execution. These are genuine gains, although it is universally acknowledged by those involved that the E-HIPC linkage has seriously degraded quality by inducing governments to rush PRSP processes in order to secure the irrevocable relief that is granted on reaching Completion Point.

So there are important positives as well as negatives, and it is not my purpose to deny that. Nevertheless, the last few pages have surely been enough to demonstrate a real danger that E-HIPC is eroding aid effectiveness. At the very least, in the face of the failure of additionality, the desirability of increased debt relief should not be taken as axiomatic. Well-targeted transfers delivered as regular aid by cost-effective means would be preferable to a further absorption of these scarce resources by means of yet more liberal debt relief. The OED report does not argue this but one does not have to read too deeply between the lines to draw that conclusion.

The question for present purposes is, who's listening and learning? With evaluations of projects and other institution-specific interventions there are, in principle at least, established channels for lesson-learning and reasons for believing that examining the

past may improve the future. This is less obviously the case with a topic like debt relief, which is 'super-macro' in orientation. Responsibility for it straddles both BWIs and a large number of creditor-donor governments. The undoubted fact that it is highly political in its motivation further lessens the prospects that evidence will govern future decisions. Despite active dissemination, it is not clear that the OED report will carry much weight, or even reach those who are influential in these matters. The Bank's own defensive, stone-walling 'Management Response' to this report is not encouraging.⁹

III -- Conditionality: another knowledge-practice gap?

Another apparent area in which there appears to be a gulf between what the known facts tell us and the continuing policies of the aid community generally and the BWIs in particular relates to the use of conditionality as a means for achieving policy change in aid-recipient countries. On the one hand, much research -- and quite a lot of Bank (but not Fund) rhetoric -- emphasizes the limitations of conditionality as an instrument for change. On the other, it is arguable that, at least among low-income indebted countries their governments find themselves expected to conform to an even wider array of policy stipulations than in the apparent heyday of conditionality in the earlier 1990s.

However, the issues here are complex and the evidential basis unsatisfactory. The issue is a large one, however, because if indeed a misplaced reliance is placed on an instrument which actually fails to deliver the safeguards which it appears to offer then it can lead to the mis-application of large amounts of public money.

*Doubts about conditionality*¹⁰

There was a veritable explosion in the use of policy conditionality during the 1980s and into the 1990s, as the World Bank became increasingly involved in structural adjustment lending, as the IMF extended the range of its own conditions from a previously fairly narrow macroeconomic focus to a much wider range of 'structural' matters, and as various other multilateral and bilateral donors increased their own use of this instrument. While there was from the beginning much controversy about the appropriateness of the design of the BWIs' conditionality, it tended to be common ground between them and their critics that these policy stipulations were carried out, for good or ill.

As the 1990s proceeded and evidence accumulated, however, a body of research grew up which cast doubt on conditionality's efficacy. It had long been well-known, for example, that a high proportion of IMF programmes broke down before the end of their (relatively brief) intended currency (Killick, 1995). Similarly, casual empiricism

⁹ 'Management response. The Heavily Indebted Poor Countries (HIPC) Debt Initiative: an OED Review.' Washington: World Bank. n.c., ca. February 2003.

¹⁰ The following borrows quite heavily from my own work on conditionality (Killick *et al*, 1998) but see also Collier *et al*, 1997, Crawford, 1997, Dollar and Svensson, 2000, and Mosley *et al*, 1995.

suggested that programmes of structural adjustment were not working well, at least in many low-income countries. My own work and that of others listed in footnote 10 began to show the limited impact of BWI programmes and to raise questions which went to the heart of the use of conditionality as a way of achieving policy change. Countries which had received large numbers of successive highly conditional credits were still rated by the BWIs as having weak, sometimes deteriorating, policies, with little apparent association between programmes and policy trends. Programmes were often poorly implemented, so that it was not surprising that they produced weak results, but non-compliance with the BWIs policy stipulations appeared rarely to be punished in any effective or consistent way. When, as was often the case, a conflict of interest was perceived locally between domestic political imperatives and BWI stipulations it was usually domestic politics that won out. Conditionality relating to governance issues was thus particularly prone to be ineffectual (Crawford, 1997, found that in only two out of 29 cases examined was donor pressure effective in inducing political change).

The use of conditionality came to be seen as in conflict with a growing consensus about the importance for effective action of local 'ownership' of chosen reforms, and as undermining the credibility, and therefore effectiveness, of the measures that were undertaken. Governments had learned that probably no more than temporary inconvenience would be visited upon them as a result of failures to implement "agreed" conditions (amply justified, in the case of the Fund, by research showing little association between past compliance and future credits -- Bird, 2002, Dreher, 2003). The BWIs (and other donors) had strong institutional imperatives to "keep the money moving", not least a desire to protect the servicing of past credits, and these were often reinforced by staff incentives within these institutions. By providing the appearance but not the reality of safeguarding against poor policy performance, over-reliance on conditionality was blamed for resulting in major misallocations and waste of public monies (Killick, 1998:168). Far better, the critics argued, would be greater insistence on local ownership and more selectivity in the choice of governments to be supported.

How did the BWIs stand in relation to the emergence of this critique? The Fund, by and large, has remained in denial (although we will see shortly that it too is beginning to show signs of acknowledging the problem) but the Bank and members of its staff have contributed quite strongly to the negative evidence. In the words of one Bank report (1995:1), "adjustment lending has mostly promoted good policies, but got weak program results." An important Bank study of Africa concluded flatly that "Conditionality as an instrument to promote reform has been a failure" (Devarajan *et al*, 2001). Another Bank report (1998) concluded that conditionality had been ineffectual where reform lacked political support and had been counterproductive in some cases. More examples could be cited but the point is that the Bank appeared to acknowledge as valid the critique of conditionality and to share it.

Institutional responses

It was reasonable, then, to expect a fairly strong movement away from reliance on

conditionality but this has not occurred. Some other donors, notably the UK and EU, have been trying to diminish such reliance, moving instead to greater selectivity and to relationships with recipient governments based far more on dialogue, ownership and partnership. There are voices within the Bank urging movement in the same direction, particularly in connection its new Poverty Reduction Support Credit (PRSC) facility. It is also the case that, following a sharp rise in the number of conditions per Bank programme during the later 1980s, there was then a decline in the average number of conditions in the latter half of the 1990s (World Bank, 2001:80). By 1998-2000 the average stood at 36, down from a peak in 1988-92 of 58. However, the biggest reductions occurred in what the Bank calls 'non-binding' conditions (from 23 to 8) with a smaller proportionate reduction in the more serious 'legally-binding' conditions (35 to 28). In any case, the 1998-2000 average was still well above the numbers prevailing in the earlier-1980s and even in the later period Bank staff regarded only 37% of all conditions as 'very relevant' to the attainment of loan objectives. There is, in short, ample scope for streamlining action by the Bank, comparable with what is occurring in the Fund (see below) but there are no strong signals that this is occurring. Moreover, the share of adjustment lending has grown to new record levels -- 64% in 2002 -- and with it the importance of conditionality (Thomas, 2002:2). There is thus a large apparent disconnect between the Bank's own evidence on conditionality and its continuing heavy reliance upon it.

What of the Fund? Perhaps instead of publicly expressing its anguish it is just quietly getting on with transforming the way it does business? Well, indeed, things are happening, in the form of a substantial 'streamlining' exercise. Introduced in 2000, this aims to reduce the number of 'structural' conditions in its programmes (which had escalated from an average of 2 per programme in 1987, 4 in 1994 and 14 in 1997-99 -- Goldstein, 2000:82) and to focus these more on actions regarded as critical to programme success and within the Fund's own core areas of expertise. Moreover, early evidence suggests that this exercise has resulted in an appreciable decline in the average number of structural conditions, but with the extent of this varying greatly from country to country (Adam and Bevan, 2001; Killick, 2002).

However, it is important to understand the limitations of the streamlining exercise. In the words of the Fund's *Annual Report* for 2001 (IMF, 2001C:45), "the main goal of streamlining was to make conditionality more efficient, effective and focussed..." It seeks to reverse the proliferation of structural conditions, calls for greater clarity in programme documents about what constitutes Fund conditionality, and seeks to ensure that, in countries where both agencies are operating, there will be a clear division of responsibility on policy matters between the Fund and the Bank. The exercise is confined exclusively to what the Fund classifies as 'structural' conditionality, with no comparable change in its traditional macroeconomic stipulations. Streamlining, in other words, is a fairly limited exercise and once we lift our eyes from the purely quantitative aspect of conditionality, it becomes clear that it does not mark any real move away from reliance on conditionality *per se*, as, indeed, the above quotation from the 2001 *Annual Report* makes clear.

A relatively new element in the situation, with potential for reduced reliance on conditionality, is the initiation of Poverty Reduction Strategy Papers (PRSP) as a focus around which the BWIs and bilateral donors can harmonise their assistance (and also their debt relief under the HIPC-II scheme). In principle, PRSPs can be viewed as an attempt to get away from old ways of doing business and to substitute these by broad-based, locally-owned strategies, in which policy commitments are self-defined by the responsible governments, subject only to “endorsement” by the BWI Boards.

There is real potential here but it is by no means clear that the move towards PRSPs has actually marked a move away from BWI-defined policy conditions. HIPC governments now have to concern themselves with further conditionality arising from the Bank’s Country Assistance Strategy papers, as well as that specific to the HIPC completion-point arrangements,¹¹ to say nothing of the stipulations of other multilateral and bilateral donors. Countries which aspire to membership of the EU also have the conditionality of that institution to worry about. HIPC conditionality alone is potentially both onerous and wide-ranging, with its content recently summarised as normally centred around macroeconomic issues, structural reforms, social sectors and ‘other poverty reduction requirements,’ especially governance and budget management issues (SPA, 2001:9). Not much is left out there! There are also reports of bilateral donors picking up structural conditions being dropped by the Fund as a result of streamlining (Debt Relief International, n.d., para 13) and of the Fund actually increasing its stipulations for actions in the governance area.

At least for HIPC countries, it seems that governments today are probably confronted by a wider range of policy stipulations than they had two or three years ago. Freedom of action should also be judged according to the importance of the various areas of policy action. Since IMF streamlining is about ensuring that conditionality is focussed on the most critical policy areas, and with both BWIs tending to cut back most heavily in the grey area of second-order benchmark, or non-binding, conditionality, here too governments may well be in a more constrained situation, rather than an improved one. It may well prove that the HIPC II-PRSP arrangements have provided a vehicle for further increasing conditionality, despite all the rhetoric of ownership.

Two other considerations rather reinforce this view. One is the prospect that the more important structural conditions dropped as a result of Fund streamlining will be taken up in Bank credits. According to a Fund staff report on initial experiences with streamlining (IMF, 2001D:17, 34), the Bank is “strengthening” its conditionality in areas, such as privatisation, health system reform and public sector reform, from which the Fund is scaling back. In a number of cases, the report states, “measures no longer

¹¹ For example, at a Commonwealth Secretariat-IMF consultation in July 2001, the Tanzanian delegate reported that his government was confronted with no less than *thirteen* specific HIPC completion point conditions, over and above those of the BWIs.

covered by Fund conditionality were incorporated as conditions by the Bank, but in others this was not the case.”

A second consideration concerns the extent of cross-conditionality between the two BWIs. There has long been a *de facto* cross-conditionality from Fund to Bank adjustment programmes but under the arrangements between the Bank and Fund concerning their PRGF and PRSC programmes the degree of cross-conditionality has been increased at the level of broad performance, although the Boards of both institutions state that cross-conditionality should *not* be applied to policy specifics within either programme, with each institution “separately accountable for its lending decisions...” (IMF-World Bank, 2001:26).

In summary, what the above adds up to is a major disjuncture between a rather wide perception that conditionality is a flawed instrument and the continuing practices of the BWIs (as well as other donor agencies). Much of the difficulty, no doubt, lies with the difficulties of persuading the Boards of both agencies of the case for moving away from conditionality and with presenting alternatives which would satisfy them. However, another source of difficulty, as Thomas (2002) has recently argued, may be that a good deal of the evidence on which the critique of conditionality was based relates to the 1980s and early-1990s, since when the record may have improved and with it the utility of using conditionality.

Is compliance improving?

Both BWIs can point to some evidence of improvement. In the case of the Bank, OED reviews of development effectiveness report satisfactory or better outcome scores from adjustment lending operations rising from around 60% in the 1980s to 86% in FY1999-00 (97% when weighted by disbursements). As regards the Fund, a recent review of the performance of conditions in 24 operations found that only 10% were not implemented, while 65% were fully implemented (Nestmann and Weder, 2002). Against this, other Fund evidence suggests that programme failure remains a large and growing problem. Recent IMF *Working Papers* provide evidence of such difficulties. Mussa and Savastino (1999: Table 2) rate as failing programmes where actual disbursements are less than half of agreed amounts and show a rising proportion of programmes failing this test over the last two decades, after an earlier period of apparently improving outcomes (percentage of programmes less than 50% disbursed):

1983-87	29%
1988-92	33%
1993-97	46%

Ivanova *et al* (2003, Table 1) similarly show that in 1992-98 only a quarter of ESAF/PRGF programmes were *not* subject to some interruption and that nearly half (45%)

experienced irreversible interruptions.¹² Similarly, the first report of the recently-created Independent Evaluation Office of the IMF, on the issue of the 'prolonged use' of Fund resources, shows, both absolutely and proportionately, that prolonged use has been a continuously, and rather steeply, rising trend since the late 1970s, continuing at least through to 2000 (IEO, 2002). This evidence is again consistent with the view that Fund conditionality is ineffectual, perhaps increasingly so.

Econometric analysis by Ivanova *et al* (2003) confirms the dominance of domestic political-economy factors in determining Fund programme success found by independent researchers, and also that neither heightened 'effort' by Fund staff nor even increased resort to prior actions could substitute for favourable political-economy conditions and exerted no significant influence on the likelihood of programme implementation (Thomas, 2003 forthcoming). Associated with this, a Fund paper on conditionality policy issues (IMF, 2001A:55) asserted the position, apparently with approval, that the primary role of BWIs "is to identify reformers, not to create them" and that "IFIs should have no illusions that their conditionality will appreciably affect the probability of reform." Indeed, the Fund's own Executive Board is on record as stating that "conditionality cannot compensate for a lack of programme ownership". These influences have led to a recent Board paper on the strengthening of country ownership in Fund programmes (IMF, 2001B).

The case for an evaluation

In brief, then, the evidential base on the more recent record with conditionality is too mixed and incomplete to permit a firm judgement on whether the earlier-identified weaknesses have been reduced and whether further action is necessary. The point here, then, is that a major research and evaluation effort, which should at least straddle both BWIs, could bring the evidence up to date and could provide answers to some of the many questions that remain unresolved:

- ◆ Given new developments, such as the HIPC-II initiative, the increased use of PRSPs and the IMF's streamlining exercise, what is the overall trend in the totality of policy conditionality? Do governments today have greater effective room for manoeuvre in the determination of policies? What difference is IMF streamlining making?
- ◆ Have there been decisive changes in institutional and staff incentives which gave priority to new lending over the implementation of past programmes?

¹² They also show an apparently more satisfactory 73% compliance with programme conditions but this figure is hard to interpret because the authors regard this figure as biased upward. For a useful very brief review of other evidence on programme effects see IMF 2001A:45-46. See also Bird, 2002, for corroboration of declining IMF programme completion rates.

- ◆ Is there evidence of more effective sanctions against non-compliance (other than that which arises from shocks) and of the application of greater selectivity in the choice of governments supported? Has the HIPC initiative made any difference in this respect, by reducing pressures for defensive lending?
- ◆ To what extent have the negotiation styles and modalities of the BWIs changed in order to foster improved relationships and greater borrower ownership of the programmes supported?

In the meantime, we appear fated to continue the present unsatisfactory mismatch between what available evidence tells us and continuing widespread reliance by BWIs and other donors on a modality of doubtful efficacy, with potentially major detrimental consequences for the effectiveness of large volumes of development assistance.

IV -- Monitoring and evaluating the donors

Background

The starting point of this final section is that the shift in favour of partnership-based programme aid has implications for the level at which evaluation work needs to be undertaken and -- in particular -- who should undertake it. Precisely because it is in support of a government's general programme of development and poverty reduction, the evaluation of this self-evidently has to be undertaken at the country-wide level. But if the notion of partnership -- and the desire to move to relationships less reliant on the dubious efficacy of conditionality -- is to be taken seriously, there are implications also for who should undertake such evaluative work. There is today increased questioning of the past pattern of aid relationships and more desire to move towards a new paradigm. The OECD-DAC has been doing much work to compare experiences and to promote greater harmonisation of donor policies and practices (e.g. see DAC, 2003). It has also begun to discuss the possibilities of evaluations led by recipient countries. There are fairly strong moves towards establishing aid relationships based on 'partnership', which implies co-equality and mutual acceptance of responsibilities. In connection with the NEPAD initiative and in collaboration with the OECD, the Economic Commission for Africa has been promoting the idea of mutual donor-recipient accountability and of joint reviews of developmental effectiveness -- another manifestation of the same desire for a changed paradigm.

At the moment, however, almost all monitoring and evaluation at the country level is by donors of recipients, raising the question, who will scrutinise the donors? The logic of partnership is that increasingly such monitoring should be jointly conducted or sponsored and include the performance of all parties, not just the recipient side. Some donors might claim that their evaluations are already joint, in the sense that ToRs,

consultants, etc must be approved by recipient governments, but most will admit in private that this is largely a formality, with little substantive meaning.

This section draws attention to an experiment conducted in Tanzania and urges wider adoption of the model employed there.¹³ So far as is known, this has been a unique experiment, although a more limited version has been employed also in Rwanda.¹⁴

The Tanzanian case

The history is as follows. In the early-to-mid-1990s there was an escalating crisis in relations between donors and Government of Tanzania (GoT), which came to a head in early-1995 at a fraught Consultative Group (CG) meeting. There was a breakdown of trust. Donors suspected large-scale mis-use of aid counterpart funds and alleged high-level corruption in the tax-collection system. For their part, the GoT saw the donors as excessively intrusive and unrealistically demanding.

In an attempt to reverse this deterioration, the Danish government decided in 1993-94 to establish an *ad hoc* 'Group of Independent Advisers' whose broad task was to explore whether and how donor-GoT relations could be improved. This met under the chairmanship of Prof. G.K. Helleiner, a well-known Canadian international economist with long-standing experience in Tanzania, and there were four other members of the group, including two Tanzanians and two other non-Tanzanians. It was important that all members were quite senior and were regarded as being independent of both GoT and donors. The group heard evidence from virtually all donors in the country, from a range of GoT ministries and agencies, and from other interested parties. The report of this group, in June 1995, was critical of both GoT and donors and presented recommendations for improving the situation. It urged major changes in relationships between the two parties and in the operational cultures of both (see Helleiner, 1995).

That might well have been that. However, the term of office of the then President of Tanzania (Mwinyi) expired in late-1995 and elections ushered in a new incumbent (Mkapa) who was determined to improve the situation and who decided to utilise the Helleiner report as a basis for doing so. The report therefore turned out to be an influential document. It led directly on to an important agreement in January 1997 between the GoT and donors to jointly set out a programme to redefine the terms of their development co-operation. The result was an agreed set of 18 points for action. These stated, among other things, that there was a need to ensure enhanced GoT

13 See also Wangwe, 2002, for a more elaborate account and a collection of key documents.

14 In 2000 the Rwandan and UK governments agreed on an independent annual monitoring of the Memorandum of Understanding between them and in 2002 the Netherlands and Swedish governments decided to adopt this model. A monitoring exercise conducted by two independent consultants in October-November 2002 therefore reviewed progress with respect to development co-operation between Rwanda and all three donor countries. (Information kindly provided by DfID, London.)

leadership in development programming, and increased transparency, accountability and efficiency in aid delivery. The elaboration of a framework for co-operation culminated in the preparation of a *Tanzania Assistance Strategy* which was finally published in 2002.

Following the GoT-donor agreement of January 1997, it was agreed that progress in implementing the agreed points would be monitored and reports on progress would be presented to meetings of the CG. It was agreed to ask Professor Helleiner to present reports to CG meetings in 1997, 1999 and 2000. At the latter meeting, it was agreed that the monitoring activity was beneficial but needed to be institutionalised. As a result, in February 2002, the GoT and donors jointly appointed an Independent Monitoring Group (IMG) to review progress in aid relationships. The IMG reported to a meeting of the country's CG in December 2002 (see IMG, 2003). Its main message was that the situation had greatly improved since 1995 but that there remained much scope for further improvements on all sides. It made a substantial number of specific recommendations for future improvements, including the following:

“Evaluations comparable with that presented in this report should be undertaken every two or three years. These should examine GoT-donor relations in the round, i.e. not be confined to donor performance. They might be given a more specific focus...

“We urge that evaluations of government-donor relations similar to our own could usefully be replicated in other aid-dependent countries and recommend that donors suggest to their respective headquarters that they promote the wider use of this type of monitoring activity in other countries.”

In terms of replicability, the following features of the IMG are significant:

- ◆ The group was again a made up of a mix of senior Tanzanian and outside members. It was chaired by a Tanzanian economist who was at that time head of the country's leading independent policy research institute. There was one other Tanzanian, a former ambassador, and three external, European-based, members, one of whom (the author of this paper) had also been a member of the original Helleiner group. The group also included a senior Ugandan member. The group had complete autonomy in deciding how it went about fulfilling its ToRs and in what it said in its report.
- ◆ There was a tendering process within Tanzania for a contract to constitute and service the group. This was won by a local think-tank organisation which provided research, logistical and administrative back-up. It was for the contractors to identify group members, although suggestions were made by various parties,

and the constitution of the group was subject to the approval of the GoT and donors.

- ◆ The activity was financed by contributions by various donors into a fund administered by the UNDP, which is the lead agency for donor co-ordination in Tanzania. The local UNDP oversaw the tendering and contractual processes. However, it did not provide secretariat services, nor was it represented on the group.
- ◆ The ToRs for the group required a mix of monitoring and evaluation, although with a greater stress on the former. They were asked both to present information on changing aid relationships and to analyse this in order to recommend ways forward.

One other important aspect of the IMG's work should be noted. There had been a tendency locally to think of it as mainly focusing on the policies and modalities of the donors. However, the IMG took a different view:

“An important principle we would urge here is that such exercises should not be confined to monitoring just one side of GoT-donor performance. It has rightly been observed that the GoT is already subjected to intense monitoring by the donors ... whereas such monitoring of the donors as occurs is much less rigorous. The question, "Who will monitor the donors?" is well put. But we would make two points about this: (a) that there is value in reviewing GoT performance independently of the donors because that is likely to yield different insights, and may reinforce the donors in some areas while protecting the government from what may be unwarranted or over-hasty judgements; and (b) that the actions of both parties are likely to be influenced by what is done (or not done) by the other, so that it is not appropriate to judge the performance of the donors except in the context of trends on the government side, and vice versa.”

Implicit in the Group's report was a desire that any future such activities should become more evaluation-based, suggesting that future groups might be given somewhat more focused ToRs, enable it to achieve greater depth by concentrating on some particular aspect, e.g. SWAs or local-level relationships.

Is this a replicable model?

Since it appears that this case is rather unique, the question arises whether independent monitoring might be replicable or whether there is something particular to the Tanzanian case which might rule it out elsewhere. It is not obvious that the latter is the case. One thing to note is that the 1995 and 2002 groups were convened in markedly different circumstances. The first instance was marked by crisis and a breakdown of trust, the task being to find ways of breaking the deadlock and of moving forward. The

2002 IMG, by contrast, was convened against the background of general perceptions that much progress had been made and that what was needed was an independent stock-taking, with pointers towards the next steps.

Since the principal new element introduced by independent monitoring is that a searchlight of enquiry is extended to the donors, it follows that there is only likely to be agreement on such monitoring where country representatives of donor agencies are willing to open themselves up in this way. There needs to be some minimum level of co-operation with the independent monitors by the local donor agencies, as well as the government, for the activity to be fruitful. However, this hurdle may not be high. The 1995 group in Tanzania was initiated and financed by just one donor (Denmark) -- and not a top-ranking one in global terms -- and was not the result of a collective donor decision. It was nonetheless not difficult for that group to obtain access to all significant donors -- they had lots of frustrations they wanted to get off their chests! The creation of the 2002 group was more a collective decision, facilitated by the existence of an active local DAC and by staff members of the UNDP, but there were dissenting voices and some opposition. Once again, however, this did not prevent the group from being able to talk to virtually everyone it wanted to question. In both cases, there were no problems in accessing government officials, although the improvements in the intervening years meant that the level and quality of intercourse with GoT officials in the second study was much the more satisfactory. Where the degree of consensus about the desirability of establishing an IMG is likely to be more important is in implementation. Its report will present recommendations addressed to both the government and donors. A reasonable level of agreement on the desirability of establishing the group in the first place, and existence of forums at which its report can be taken forward, will enhance the possibilities of follow-up action.

One difficulty in the way of wide adoption of the Tanzanian model is the absence of obvious incentives to do so on the donor side. They are, after all, volunteering to open themselves up to critical independent investigation and their representatives on the ground are generally all too well aware that there are often plenty of grounds for criticism. In overcoming the resulting natural reluctance, perhaps the most important ingredient in both Tanzanian studies -- and this is something that is potentially present in all aid-receiving countries -- was the existence of individuals minded and positioned to push forward the idea of an independent group. The Danish Ambassador was a crucial player in the mid-'90s; Professor Helleiner and UNDP staff were crucial moving spirits in setting up the 2002 group. Basically, all it needed was a few well-placed individuals from the donor and government sides. There is, as suggested earlier, greater openness now to a search for more satisfactory aid relationships; independent monitoring and evaluation can be used as part of a new model. Others may be unenthusiastic, even hostile, but are likely to want to have their voices heard when a group is actually formed.

What are the key ingredients of a successful group? The following suggest themselves:

- ◆ Its members -- and particularly its chairperson -- should be sufficiently senior and experienced to command respect. Above all, they should be perceived as independent of any particular interest. It is important that it should contain a good balance as between nationals of the country in question and of the outside world. The idea adopted in the second Tanzanian study, of involving a senior official from another country within the same region is also an attractive one.
- ◆ Its financing, so far as possible, should be structured so as to protect the group from suspicions that they might be influenced by anxiety not to offend those providing the money. Similarly, it is better that the group is serviced outside of the government or any donor. The arrangements for the 2002 group in Tanzania were excellent in this respect.
- ◆ The terms of reference of the group should not be too restrictive. They should allow flexibility for the group's enquiries to roam as appears necessary in the light of the information that comes to them. Addition of wording along the lines of "...and such other subjects as the group may deem desirable" is what is needed.
- ◆ There should preferably be mechanisms in place that allow the group's recommendations to be considered and acted upon. In Tanzania, the 1995 group was lucky because of the emergence shortly afterwards of a new reform-minded political leadership. By the time of the 2002 group, well-organised avenues of dialogue and co-ordination -- most notably a Tanzanian Ministry of Finance with a lively interest in these matters which had already developed a national assistance strategy and an active and well-led local DAC -- had evolved to take forward the group's recommendations (although it is too early yet to know how much follow-through there will actually be). Again, the importance of the activist voice should not be discounted.

To sum up, the above account strongly suggests that the model of the Tanzanian IMG could readily be adopted in other aid-dependent countries. There is nothing uniquely Tanzanian about it. In the context of the conduct of evaluation studies to promote greater aid effectiveness, the above account points to the desirability of taking the IMG model further as a modality well suited to aid relationships which aspire to break out of the mould of the past and to develop in the directions of partnership, mutual accountability and harmonisation.

(end)

t.killick@odi.org.uk

July 2003

Works cited

- Adam, C.S. and D.L. Bevan (2001), 'PRGF stocktaking on behalf of DfID.' Oxford University, Department of Economics (processed draft). November.
- Bird, Graham, 'The completion rate of IMF programmes.' *World Economy*. 25(6), June, 2002.
- Bird, Graham and Alistair Milne 'Debt Relief for Low Income Countries: is it Effective and Efficient?' *World Economy* 26(1), January 2003.
- Birdsall, Nancy and John Williamson, *Delivering on Debt Relief* (Washington: Center for Global Development and Institute of International Economics, April 2002).
- Brown, A., F. Naschold, T. Conway and A. Fozzard, 'Aid transaction costs in Viet Nam.' London: Overseas Development Institute. December 2000 (processed).
- Bulić, Aleš and A. Javier Hamann, 'How volatile and unpredictable are aid flows, and what are the policy implications?' *IMF Working Paper WP/01/167*. Washington: IMF. October 2001.
- Castro-Leal, Florencia, Julia Dayton, Lionel Demery and Kalpana Mehra (1999) "Public social spending in Africa: Do the poor benefit?" *World Bank Research Observer*, 14(1); February.
- Collier, P., P. Guillaumont, S. Guillaumont and J.W. Gunning, 'Redesigning conditionality.' *World Development*, 25(9), September, 1997.
- Crawford, Gordon, 'Foreign aid and political conditionality: issues of effectiveness and consistency.' *Democratization*, 4(3), 1997.
- Dagdeviren, Hulya and John Weeks, 'How much poverty could HIPC reduce?' (p.17). Processed, August 2001. Available at www.wider.unu.org.
- Debt Relief International (n.d., 2001?), 'Reviewing PRSPs: the views of HIPC ministers and PRSP co-ordinators .' London: DRI (processed).
- Department for International Development (DfID), *Eliminating World Poverty: Making Globalisation Work for the Poor*. London: DfID. December 2000.
- Devarajan, Shanta, David Dollar and Torgny Holmgren, 'Aid and reform in Africa.' Washington: World Bank, 2001.
- Development Assistance Committee of the OECD (DAC), *Harmonising Donor Practices for Effective Aid Delivery*. Paris: OECD. 2003 (forthcoming).

Dollar, David and Jakob Svensson (2000), 'What explains the success or failure of structural adjustment programmes?' *Economic Journal* 110, October.

Dreher, Axel, 'The influence of elections on IMF programme interruptions.' *Journal of Development Studies*. Forthcoming, 2003.

Easterly, William, 'How did highly indebted countries become highly indebted? Reviewing two decades of debt relief' World Bank *Working Paper* 2225. Washington: World Bank. 2001.

Foster, M., A. Brown and T. Conway, *Sector-wide approaches for health development: a review of experience*. Geneva: WHO. 2000.

Goldstein, Morris (2000), "IMF structural conditionality: how much is too much?" Washington: Institute for International Economics (processed). October.

Gunter, Bernhard, 'Does the HIPC initiative achieve its goal of debt sustainability?' (Processed, 2001, available at www.wider.unu.edu)

Helleiner, G.K. *et al*, *Report of the Group of Independent Advisers on Development Co-operation Issues between Tanzania and its Aid Donors* ('the Helleiner report'). Copenhagen: Royal Danish Ministry of Foreign Affairs. June 1995.

Independent Evaluation Office, of the IMF (IEO). 'Evaluation of prolonged use of IMF resources.' Washington: IMF. September 2002.

Independent Monitoring Group, *Enhancing Aid Relationships in Tanzania* ('the IMG report'). Dar es Salaam: Economic and Social Research Foundation. Final draft, March 2003.

International Monetary Fund papers:

'Conditionality in Fund-supported programs -- policy issues.' IMF (PDR) (processed), February 20, 2001 (A)

'Strengthening country ownership of Fund-supported programs.' IMF (PDR) (processed), SM/01/340. November 13, 2001 (B).

Annual Report, 2001. Washington: IMF, 2001 (C)

'Streamlining Structural Conditionality -- Review of Initial Experience.' Ref: SM/01/219 (processed). July 12, 2001 (D)

IMF-World Bank (joint), 'Strengthening IMF-World Bank Collaboration on Country Programs and Conditionality.' Washington: IMF-World Bank (processed), August, 2001.

Ivanova, A., W. Mayer, A. Mourmouras and G. Anayiotos, 'What determines the success or failure of Fund-supported programs?' Washington: *IMF Working Paper WWP/03/08*. January 2003.

Killick, Tony, *IMF Programmes in Developing Countries: Design and Impact*. London and New York: Routledge and Overseas Development Institute, 1995.

Killick, Tony, *Aid and the Political Economy of Policy Change*. London: Routledge and ODI, 1998.

Killick, Tony, 'The 'streamlining' of IMF conditionality: aspirations, reality and repercussions.' Report for Department for International Development, London. April 2002 (processed).

Killick, Tony, 'Helleiner on Africa in the global economy'. *Canadian Journal of African Studies*. 2003 (forthcoming).

Lawson, A., D. Booth, A. Harding, D. Hoole and F. Naschold, 'General budget support evaluability study, Phase I. Final synthesis report' (a report to DfID). Oxford Policy Management and Overseas Development Institute. December 2002 (processed).

Mosley, Paul, Jane Harrigan and John Toye *Aid and Power: The World Bank and Policy-based Lending*, Volume 1. London: Routledge, 2nd edition, 1995.

Mussa, Michael and Miguel Savastano (1999), 'The IMF approach to economic stabilization.' Washington: IMF *Working Paper*, WP/99/104. July.

Nestmann, Thorsten and Beatrice Weder, 'The effectiveness of international aid and debt relief: a selective review of the literature.' Paper prepared for the 5th Limburg seminar on financing and development 2002, processed.

Operations Evaluation Department (OED), *The Drive to Partnership: Aid Co-ordination and the World Bank*. Washington: OED, World Bank; 1999.

Operations Evaluation Department (OED), *The Heavily Indebted Poor Countries (HIPC) Debt Initiative: An OED Review*. Washington: OED, World Bank. February 2003. Report No. 25160.

Overseas Development Institute (ODI), 'Can we attain the Millennium Development Goals in education and health through public expenditure and aid?' London: *ODI Briefing Paper*, March 2003.

Pattillo, Catherine, Hélène Poirson and Luca Ricci, 'External debt and growth.' *IMF Working Paper* WP/02/69. Washington: IMF. April 2002.

Ranis, Gustav and Frances Stewart, 'The debt relief initiative for poor countries: good news for the poor?' *World Economics*, 12(3), July-September, 2001.

SPA Task Team on Contractual Relationships and Selectivity (2001), 'Comparative review of I-PRSP targets and conditionalities for HIPC Completion Point.' Brussels: European Commission (DEV/B/2/FCS D), processed. October.

Thomas, Allun. 'Prior actions -- true repentance? An evaluation based on IMF programs over the 1992-99 period.' Washington: *IMF Working Paper* (forthcoming, cited by Ivanova *et al*, *op cit*).

Thomas, M.A., 'Can the World Bank enforce its own conditions? Washington: World Learning. n.d., ca. 2002.

Wangwe, Samuel M. (ed), *NEPAD at Country Level: Aid Relationships in Tanzania*. Dar es Salaam: Mkuki na Nyota Publishers. 2002.

White, Howard and Tony Killick with S. Kayizzi-Mugera and M-A Savane, *African Poverty at the Millennium* (Washington: World Bank, 2001).

White, Howard and Geske Dijkstra, *Programme Aid and Development: Beyond Conditionality*. London and New York: Routledge. 2003.

World Bank publications:

'Higher impact adjustment lending'. Report of a Working Group to SPA Plenary. Washington: World Bank. October, 1995.

Assessing Aid: What Works, What Doesn't and Why. Washington: World Bank. 1998.

Adjustment Lending Retrospective: Final Report. Washington: World Bank, Operations Policy and Country Services Department. June 15, (2001 (A)

Global Development Finance, 2001 (B), pp.260-61