



May 5-6, 2005 Cape Town, South Africa

BUDGET SUPPORT, CONDITIONALITY AND IMPACT EVALUATION

Jan Willem Gunning
Free University, Amsterdam,
AIID, CSAE, EUDN and Tinbergen Institute
Department of Economics
Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

E-mail jgunning@feweb.vu.nl

Phone: 31 (20)598 6141



The World Bank

Budget Support, Conditionality and Impact Evaluation

Jan Willem Gunning

1. Introduction

The perceived failure of conditionality has led to proposals to tie aid to results achieved rather than to policy changes. Such results-based lending has now started to be implemented. In this paper we review some of the implications for the relationship between donors and the recipients of aid.

The structure of the paper is as follows. The next section sketches the background for results-based aid. We note that in an extreme form of selectivity aid allocation is based on measured performance but this is not expected to change the incentives of recipient governments. However, most donors do expect their allocation decisions to have incentive effects. Section 3 therefore considers aid as a contract. We consider the question how performance targets should be set and whether governments should be compensated for adverse events beyond their control. In section 4 we discuss the EU experience with results-based aid. Section 5 considers the effect of shocks and the possibility of introducing an insurance element in aid contracts. Here we address the question whether making aid results-based introduces undesirable volatility in aid flows. In section 6 we discuss the implications of the switch to budget support and the growing reliance on performance indicators for the evaluation of aid. Section 7 concludes.

2. The Case for Results-Based Aid

The idea for results-based aid grew out of the critique of conditionality. Donors had, of course, always attached conditions to aid, usually at the level of individual projects. Sometimes the objective was to promote donor interests (e.g. by requiring that bidding for procurement for a project would be restricted to firms in the donor's country). More commonly, the donor was convinced that imposing conditions would raise the project's returns and would therefore be in the interest of the recipient government. Such paternalism might be a source of friction since, obviously the two parties could disagree on what best served the government's interests.¹

¹ Collier *et al.* (1997).

Ex ante conditionality at the project level has been criticised on many different grounds. The fungibility critique is the most important one. Under fungibility the donor deludes himself: what his aid makes possible is the government's marginal project, which may be very different from the project the donor finances. In this situation "what you see is *not* what you get" and conditionality makes little sense.²

The importance of aid as project finance receded in the 1980s when donors tried to use structural adjustment lending to change the policy environment in developing countries. Conditionality was increasingly applied at the national level (e.g. in terms of trade policy, fiscal reform or privatisation) rather than at the level of individual projects. There now exists a substantial empirical literature (using case studies or growth regressions) on such "policy lending". While these studies differ in many ways there is very broad agreement on the conclusion that the type of conditionality used under structural adjustment did not work.

In some cases (as in Zambia under Kaunda) the reforms required by donors simply were not implemented, in spite of conditionality. More commonly, they were implemented, but as a result of a domestic policy debate rather than donor pressure. Vietnam and Uganda are examples of this. Finally, there are many instances (such as the famous case of food marketing liberalisation in Kenya) where conditionality did succeed in bringing about reforms but these were not maintained.³

The experience with structural lending revealed the ineffectiveness of *ex ante* conditionality where aid was contingent on *promises* of reform. In the 1990s *ex post* conditionality was proposed as an alternative. Under *ex post* conditionality (sometimes called results-based or performance-based aid) the amount of aid depends in a previously agreed way on progress realised over a given period and measured by an agreed set of indicators.⁴

The idea could easily have remained a pipe dream of academics. That it did not owes much to the conjunction of two developments. First, the *Assessing Aid* report (World Bank, 1998) became extraordinarily influential in spreading the message of the failure of conditionality very widely. Secondly, rational choice models of government behaviour in response to aid contributed to a growing awareness that the effectiveness of aid depended crucially on the incentives for governments to use aid in desirable ways.

A well-known example is the political economy model of Adam and O'Connell (1999). Here private agents choose a level of savings and allocate their savings between an informal sector (where they escape taxation) and a formal sector (which is taxed). By its

² World Bank (1998).

³ See e.g. Gunning (2001) for a review of this literature. Reform of maize marketing was a World Bank condition. The Kenyan government agreed, implemented the reform and subsequently reversed the liberalisation. Eventually the reform was implemented and reversed seven times. The episode has been cruelly but accurately summarised by the saying that the Kenyan government sold the same reform to the World Bank (indeed to the same Bank vice-president) seven times in a row.

⁴ For example Collier *et al.* (1997).

choice of a rate of taxation the government determines the growth of the economy, through the incentives provided to private agents for investing in the formal sector. The government serves the interests of a subset of the population: it uses the aid it receives and its tax revenue to finance an amount of government expenditure (which is fixed in this model) and transfers all remaining resources to the favoured subset of the population.⁵

The size of the subset determines the nature of the state. If the subset is very small the government has strong incentives to tax. At the margin all additional tax revenue accrues to the favoured group while the cost of taxation (lower growth) is borne by the entire economy. The cost-benefit ratio becomes less favourable for taxation if the subset grows. For that implies that there is less scope for letting the excluded group bear the costs. At some critical level the state will be “sufficiently representative”. It will still serve the interests of a subset but it now is optimal to reduce taxes and thereby promote growth.

It is easily seen that these two regimes have radically different implication for the effects of aid. If the regime is narrowly based (*i.e.* if the favoured subset is very small) then aid will be used entirely to increase transfers to the favoured group. Aid is then ineffective in raising growth. If the state is sufficiently representative, however, it will act as a “developmental state”: at the margin aid is then used to promote growth rather than to increase transfers.

Such models help to explain the increasing support for selectivity in the allocation of aid. Apparently there are countries in which aid cannot effective, simply because the government has no incentive to use it in a way which would make it so. Since the evidence from the conditionality literature suggests that donors cannot “buy” reform, donors should simply accept that they have no role in such countries. Selectivity then implies that aid is allocated to countries (and only to those) where it can be effective.

The selectivity approach does not rely on incentive effects. However, selectivity would of course be reinforced if recipient governments strive to satisfy the criteria which would qualify them for aid. In addition, if the amount of aid is a function of performance and this is know to the government then it may have an incentive to improve its performance, as perceived by the donor. Such incentive effects are central in the discussion on the design of the “aid contract.”

3. Aid as a Contract

While the language of “cooperation” and “partnership” suggests that donors and recipient governments pursue common goals, the reality of aid relationships is, of course, often very different. Once it is recognised that the two parties - donors and recipient governments - may have very different aims it makes sense to see their relationship in contractual terms.

⁵ One could interpret the fixed amount of government expenditure (other than transfers) as a reflection of the incentive contract of the excluded part of the population. They will consent to the regime’s policies provided they get a particular level of public goods provision.

The challenge is to design what economists call an incentive-compatible contract. This is an arrangement where one party (the principal) has imperfect control over the other party (the agent) but the terms of the contract give the agent an incentive to behave in a way which promotes the interests of the principal.⁶ Contract theory focuses on situations in which the interests of the two parties diverge and where there is asymmetric information so that the principal can only imperfectly observe the actions of the agent. In the political economy literature on aid this principal-agent framework is applied with the donor as the principal and the government as the agent.

As the Adam-O'Connell model illustrates very clearly, aid will be wasted (from the donor's point of view) when it is unconditional and the government has no incentive to promote growth and reduce poverty or to pursue other objectives cherished by donors.

Selectivity in aid allocation can be seen as a radical response to this fact: if aid is likely to be wasted in such environments then it should be allocated exclusively to countries where governments have an incentive to use aid in a poverty reducing way.⁷ Since donors have been singularly unsuccessful in "buying policy reform" there is no point in trying to use aid to promote policy change. In this view aid is therefore tied to success, not as an incentive for governments to choose policies which might generate success, but simply to ensure that aid is not wasted but allocated to countries where it can be productive.

A less extreme view would leave some room for an incentive effect of aid. Donors may succeed in influencing government behaviour if they tie aid to results *and if that link is perceived as credible*. The emphasis on incentives implies that the relationship between donor and recipient government is indeed seen as a contract.

Performance-based aid (*ex post* conditionality in the jargon) can be seen as a practical way of implementing this moderate version of selectivity. The donor provides aid in the form of budget support and the amount of aid is increasing in performance as measured by a set of agreed indicators. Clearly, the framework of contract theory is very relevant for the analysis of such budget support. Here we briefly consider four issues of contract design for budget support: completeness of the contract; the choice of target levels for the various indicators; procedures for audit and evaluation; and risk sharing between the contract partners.

First, consider the completeness of contracts. Donors and governments typically set multiple targets. This raises the question to what extent aid will be reduced if one or more of the targets are not met. In *ex ante* conditionality a similar situation arose: in principle all conditions had to be met and failure to meet any one of them would therefore in theory result in *all* aid being withheld. Such a drastic response would seem so

⁶ Contract theory has become very influential in development economics. For a textbook example see Ray (1998), e.g. ch. 15. Azam and Laffont (2003) is an example of the use of contract theory to analyse conditionality.

⁷ The most influential statement of the case for selectivity is probably the *Assessing Aid* report, World Bank (1998).

inappropriate that this all-or-nothing aspect of *ex ante* conditionality came to be seen as incredible.

A similar issue arises for *ex post* (performance-based) conditionality: if all targets are to be met then a single failure would jeopardise the entire aid budget. As before, this would make the aid contract incredible. Alternatively, the contract would have to specify a formula, indicating how much aid is to be “earned” with progress on each of the indicators. Such a formula would indicate trade-offs: superior progress in one dimension would compensate disappointing outcomes in another dimension.

So far many donors (with the notable exception of the European Union, discussed in section 4) have been rather reluctant to specify such trade-offs and commit themselves to a “mechanistic” formula. As a result the early experience with performance-based lending was characterised by remarkable vagueness: targets were specified but it was unclear how much money was riding on them. Obviously, such an incomplete contract cannot be effective in term of incentives.⁸

Secondly, consider how the targets set in the contract can be made realistic. Where there is uncertainty over what is feasible there may well be a danger that the recipient government will try to set unambitious targets. The EU recognises that if one country sets a modest target it may get more aid than another country which has made more progress but which had set more ambitious targets.⁹ The incentive to play safe is not just a theoretical possibility. In Uganda in 2001 when targets in the education sector were missed the government reacted by lowering targets.¹⁰ Making targets realistic would require some international comparability.

Collier *et al.* (1997) proposed a regression-based approach to set growth targets. A growth regression with various controls for initial conditions and geographical growth determinants such as landlockedness could be used to derive an estimate of a “normal” growth rate, in the sense of the expected value for a country with a particular set of characteristics. Targets could then be set in such a way that the “normal” amount of aid would be disbursed if this average performance were realised, with revisions upwards or downwards for better or worse performance.

Clearly, this approach can be extended to other performance indicators. The target would be seen as fair since it would incorporate all available information on the difficulty of progress in a particular set of circumstances. It would be difficult to argue that an international average, corrected for observable inter-country differences, sets too high a standard. Such an approach would remove the incentives for donors and governments to set low targets so as to avoid acrimonious debate when the targets are not met. In addition, such a system might introduce an element of international competition. This

⁸ Cf. Collier *et al.* (1997) and Adam and Gunning (2002).

⁹ European Commission (2005), p. 40.

¹⁰ Adam and Gunning (2002). The EC somewhat downplays the issue, stating that “Revisions of target levels made so far do not seem unduly lacking in ambition” (European Commission, 2005, p. 43).

could well provide a powerful incentive as illustrated by the effect of India-China comparisons in the debate on economic policy in India prior to the liberalisation.

Thirdly, if aid is tied to measured performance then a key question is who collects those data. Obviously, measured outcomes may be inconvenient, both to the government and to donors. Leaving the responsibility for verification to the contract partners therefore raises a moral hazard issue. As yet, there are few indications that data have been withheld or altered but the incentives to do so are clearly present. There seems to be a strong case for independent data collection: once indicators are agreed the audit function could be assigned to independent consultants or researchers. Just as donors have come to recognise the need for independence in evaluation, they should now recognise the case for independence in data collection.

Finally, consider the perennial question whether results-based aid allocations should be corrected for circumstances over which the government has little or no control.¹¹ For example, suppose the donor has tied aid disbursements to reductions in infant mortality and infant mortality rises as a result of a drought. Should aid be cut because the infant mortality target has not been reached? If so, by how much? Or, alternatively, should donors in fact offer insurance, compensating for the drought by raising aid above the level the government would otherwise be entitled to?¹²

The answer depends on whether the government can mitigate the impact of the drought on infant mortality and, in addition, whether its efforts to do so can be reliably monitored by the donor. Very likely, there is considerable scope for government action (e.g. through targeted drought relief, food storage programs and development of credit and insurance markets). But it is also likely that such government actions can at best be only imperfectly monitored by the donor.

In the debate on aid effectiveness one can distinguish two radically different solutions to problems of this kind. One school argues that the extent to which the government has affected the outcome is a matter for *ex post* negotiation between the two parties. In effect this amounts to abandoning results-based aid (*ex post* conditionality). For under such an arrangement there is no contract *ex ante* specifying state-contingent aid disbursements. Since the terms of the contract are left unspecified they can have no incentive effects. In effect this school argues for a return to *ex ante* conditionality. Since there now is overwhelming evidence that such conditionality is not effective, this solution is simply not credible. However, it enjoys considerable popularity.

What is the alternative? If one analyses the problem in terms of contract theory it is readily seen that the situation is analytically equivalent to the case for partial insurance. The argument is straightforward.¹³ Suppose the outcome (in this case: infant mortality)

¹¹ See Collier *et al.* (1997) and European Commission (2005, p. 49).

¹² There is an old literature on the use of aid to compensate developing countries for trade shocks, through mechanisms such as the European Stabex program. See *e.g.* Collier *et al.* (1999). Guillaumont and Chauvet (2001) revived the case for aid as insurance, in the context of the debate on criteria for the allocation of aid across countries. Cf. Gunning (2001, 2004).

¹³ See *e.g.* Ray (1998), pp. 602-605, for a formal treatment.

depends both on exogenous shocks and on efforts of the government. In particular: the government is able to reduce (but not to zero) the probability of a bad outcome (and hence increase the probability of a good outcome). However, the donor is unable to monitor the government's efforts. Hence the donor observes outcomes but is unable to determine to what extent these are due to exogenous events beyond the government's control or to the government's actions (or lack thereof) in response to the *possibility* of events such as a drought.

Since the donor cannot observe the government's efforts the contract can obviously only be written in terms of outcomes. Both parties recognise, of course, that a bad outcome may not be the government's fault, but they understand that there is no alternative to writing the contract in terms of outcomes. How much aid will be disbursed is therefore mechanistically determined by the observed outcome. The outcome would amount to full insurance if aid were independent of the outcome. The donor would then in effect fully compensate the government for a poor outcome. Clearly, this would remove any incentive for the government to exert itself so as to reduce the probability of a bad outcome.

Under an efficient contract the government will make that effort in spite of the fact that it cannot be observed. Partial insurance achieves this by imposing risk sharing on the two parties. Technically, the contract must satisfy the *incentive constraint*. This means that when the government chooses to allocate resources to, say, programs affecting child health it thereby reduces the probability of high infant mortality sufficiently to make this worth its while, through the increase in *expected* aid. Under such a contract the government will receive *less* aid (compared to a simple results-based aid program) when the results are good and *more* aid when the measured results are poor. In this sense the aid involves an element of insurance.

The first point to note is that this insurance is partial. Complete insurance is not incentive compatible under asymmetric information: the government would become passive. This is relevant for the case of budget support. It is often suggested that results-based budget support should be amended in such a way that governments are *fully* compensated for events beyond their control. As the example of drought nicely illustrates, the issue is not whether an event is beyond the control of the government. When a drought occurs that fact can, of course, easily be established. The issue is rather to what extent a performance indicator such as infant mortality is affected by such an event. Clearly, this is a very difficult question and in this area the perfect may well be the enemy of the good.

However, there is some scope here for advancing beyond pure guesswork. For example, cross-country regressions can be used to estimate the effect of various easily observable variables (such as the occurrence of a drought) on infant mortality.¹⁴ This goes some way towards estimating the impact of government actions. Alternatively, and more convincingly in my view, one can use microstudies (impact evaluations) to estimate this effect more directly. This information can then be used as input into the design of the aid contract.

¹⁴ This is the approach advocated by Collier *et al.* (1997) to correct for, say, the impact of trade shocks.

As the example shows, the design of aid contracts involves two information problems. One can be called model uncertainty: we do not know the effect of the government's actions on the chosen performance indicators. The other information problem is that the government's actions themselves are not known, at least to the donor: this is the standard case of informational asymmetry in contract theory.

The second point to note is that some of the risk is borne by the donor. That this is efficient seems to be insufficiently appreciated by the donor community. Understandably donors want to keep aid contracts as simple as possible. They therefore shy away from making these contracts state-contingent. Since in practice this leaves all risk with the government such an arrangement gives the government the maximum incentive to take measures which increase the probability of good outcomes. However, the arrangement may hold little attraction for a risk averse government.

Where recipient governments typically have very limited scope to deal effectively with risk, refusing to extend insurance is unlikely to be an effective aid policy. To take an example, the empirical literature on the effect of trade shocks shows that positive shocks contribute very little to growth in developing countries while negative shocks substantially reduce growth. This suggests that there is a case for aid as insurance.¹⁵

The donor could use the regression evidence to derive an estimate of the effect of a particular event (a drought) on an indicator (an increase in infant mortality by x). It would like to give the government incentive to take measures which reduce the impact. Very likely it lacks the information (e.g. how much compensation is needed to induce the government to take additional measures) needed to design an efficient partial insurance contract. What it can do is to offer partial compensation (a fraction of x) in the aid contract contingent on the occurrence of a drought.

The question whether performance-based aid should be amended so as to offer an element of insurance is one to which we return in section 5.

4. Implementing Results-Based Aid: Design and Experience

Results-based aid was long seen as an obsession of impractical economists. That the idea was implemented owes much to the European Commission.¹⁶

The key characteristic of European budget support to ACP countries is that part of it is a fixed tranche, part a variable tranche. Payment of the fixed tranche is tied to satisfactory macroeconomic performance and retains the all-or-nothing character which made traditional *ex ante* aid incredible: in principle a failure on any aspect of macroeconomic

¹⁵ See Collier and Gunning (1999) on the trade shocks evidence and Collier and Dehn (2002) on the impact of aid and shocks on growth.

¹⁶ See Adam and Gunning (2002) and Adam *et al.* (2004) for early assessments of the EU's performance-based lending in four African countries. For a more recent and much more comprehensive assessment see European Commission (2005).

performance would justify withholding the entire tranche. The obvious comparison is with a nuclear threat: if the only options are doing nothing or imposing enormous damage the threat to “go nuclear” for any, possibly minor, infringement is not credible. At present the variable tranche is still only fairly modest, accounting for about 35% of the total amount.

In its review of 34 programmes the EC found that the percentage of the variable tranche disbursed ranged from 65 to 85%.

The Commission distinguishes four types of performance indicators: those measuring input, output, outcome and impact. Input measures the resources provided, e.g. the budget allocation for primary schooling. Output measures the direct results of the activities financed. For example, output in the education sector can be measured by the number of schools built, the number of schoolbooks delivered or the number of qualified teachers. Outcome is vaguely defined as “results at the level of beneficiaries”; examples include school enrolment. Impact measures ultimate objectives, e.g. poverty, literacy or morbidity.

The EU approach is gradually gaining wider support. For example, both the Swedes and the Swiss are experimenting with variable tranches tied to performance in Mozambique. It may be noted that as more donors shift to budget support and there is, through better donor coordination, increased agreement on the indicators to be used to measure success, the all-or-nothing alternative to a graduated system becomes increasingly untenable. In the limit, when donors fully agree on the indicators, either all donors would cut off aid or none would. Coordination therefore very much worsens the “nuclear threat” aspect of all-or-nothing systems. Hence such aid systems become less credible as donor coordination improves. As donors become aware of this growing credibility problem they are more likely to adopt graduated response systems.¹⁷

In practice this means that performance is measured on the basis of a small number of indicators; that these are scored (0, 0.5 or 1); and that a weighted average of the scores is calculated. This then determines the fraction of the variable tranche which is disbursed.

A number of aspects of the European system are worth noting. First, the system is entirely transparent: the government knows how much it will receive, depending on the performance indicators. The link between results and disbursements is, in the phrase of the EU’s critics, mechanistic. There has been an unfortunate episode where “grossing up” was applied (scores were multiplied by 1.25 after the fact to so as to be able to disburse higher amounts) but this was quickly abandoned.¹⁸

Secondly, the indicators used are measures of output or outcome, not of impact. Indeed, the Commission insists that the indicators used for the variable tranche need to be measured annually.¹⁹

¹⁷ European Commission (2005), p. 47.

¹⁸ European Commission (2005), p. 39.

¹⁹ European Commission (2005), p. 7.

Thirdly, about two-thirds of the aid provided belongs to the fixed tranche and is tied to satisfactory macroeconomic performance. My interpretation is that the Commission is gradually moving from one system to another and that eventually the entire aid programme will be of the variable tranche type. However, the present hybrid is difficult to understand. Since the fixed tranche has the old all-or-nothing character the macroeconomic conditions are apparently “show-stopping” conditions. However, in that case the entire program (rather than just the fixed tranche part of it) should cease if the conditions are not satisfied. More importantly, as in our earlier examples for Uganda and Zimbabwe, the “show-stopping” conditions donors have in mind are typically not in the area of macroeconomics. This suggests that recipient governments are confronted with a complicated and potentially confusing set of signals. There are circumstances (apparently not clearly spelled out) which will stop the entire program; there are macro-economic events which might stop the fixed tranche but where the credibility of that threat suffers from the nuclear deterrent problem: it is difficult to believe that the donor would choose such a drastic approach if only some of those conditions were violated; and there is a category of events which affect the performance indicators with precisely known consequences for the disbursement of the variable tranche.

An additional concern is that the macro-economic conditions are often rather vague, leaving much room for interpretation, disagreement or negotiation.²⁰ For Malawi in 1999 what was required was “adherence to the structural adjustment programme’s macroeconomic framework and ESAF benchmarks”. For Uganda in 2000 the Commission required a “satisfactory macro-economic background”, for Benin in 2003 “conclusions positives a l’issue des revues de la FRPC” and for Rwanda in the same year “le gouvernement doit poursuivre avec le FMI son programme de réformes macro-économiques”.

Adam and Gunning (2002) in their preliminary assessment of the experience with the EU’s system in Uganda concluded that the aid contract had as yet changed little.

5. Shocks, Results and Predictability

A key concern in the debate on aid effectiveness is the predictability of resource flows. Particularly in African countries - where aid often finances a very large part of public expenditure - there is a concern that effective management of public expenditure requires long run commitments. It is sometimes suggested that this may be difficult to reconcile with results-based lending: conceivably aid flows can suddenly diminish if targets are not met.

This point is not well taken: it confuses *volatility* with *predictability*. It is true that under results-based lending aid flows can be volatile. This would arise if the extent to which targets were reached would vary substantially from one year to another. However, to the extent the government can influence the outcomes, aid flows are fully predictable. This is, of course, the very idea of results-based lending: the contract has an incentive effect

²⁰ European Commission (2005), Table 3a.

only if the government knows exactly how much aid it will receive as a function of the outcomes as measured by an agreed set of performance indicators. In these circumstances aid flows can diminish but this would not come as a surprise. Obviously, this is only part of the story: outcomes are also affected by events which are not under the control of the government and this does introduce an element of unpredictability.

However, it is a mistake to view this unpredictability as a disadvantage of results-based aid. Presumably the counterfactual would be a situation in which aid was stable (not being tied to possibly volatile outcomes) whereas outcomes were not. Given the choice the government would *prefer* the unstable aid flows under the results-based aid contract if that were designed as a partial insurance contract, as discussed in the previous section. Indeed the point of the incentive constraint, a key requirement of such a contract is that the government has no incentive to reject the insurance-inclusive contract.

Hence it is misleading to point to the implication of results-based aid that resource flows may be less stable. The point is rather that under this system the government will face *less* risk than under stable aid flows. A risk-averse government will therefore *prefer* the unstable aid flows implied by the partial insurance contract. Once it is recognised that the aid contract involves insurance the apparent contradiction is easily resolved: the contract enables the government to share some of the risk it faces with the donor.

However, the EU position is quite different. In response to the question “Should the government be penalised for something for which it is not solely responsible?” the Commission makes two points.²¹ The first is that while the chosen performance indicators “are not directly within Government’s control, they are measures over which the government has very large leverage, for good or ill: if primary enrolment is low, for example, while it is true that Government cannot (and probably should not) force children into schools, enrolment is probably low because schooling is unaffordable, inaccessible, or of such low quality that parents see no point in sending their children to schools. All of these are factors within Government’s control .. Hence in general the kinds of indicators used are readily susceptible to Government action.”

This is obviously true for outcome indicators such as school enrolment, but it is not true for impact indicators. These measure progress towards higher level targets such as the Millenium Development Goals for poverty or under-5 child mortality. Government control over such targets is only partial. This poses a dilemma. Either results-based lending is gradually shifted toward reliance on impact indicators but then the claim that the measures are within the government’s control is no longer valid. Or, the aid contract continues to be defined in terms of measures under government control but then the contract provides no incentives to strive for higher-level development goals. Before long this dilemma will have to be faced.

The Commission summarises its position on general (as opposed to sector-specific) budget support as follows:²²

²¹ European Commission (2005), p. 49.

²² European Commission (2005), p. 50.

“.. we judge it appropriate to use a small number of results indicators, at outcome level, recognising that this strikes the best balance between what is swift-changing enough to be meaningful annually, close enough to Government control to be relevant to public decision-making, but sufficiently synthetic to capture a wide range of actions sufficiently closely linked to actual outcomes (indeed, often being desired outcomes themselves) to be strong measures of poverty reduction itself.”

Note that here it is taken as self-evident that indicators should be “swift-changing enough to be meaningful annually.” At the same time these will be “strong measures of poverty reduction itself”. This combination is neither feasible nor desirable. It is not feasible, because poverty reduction is typically a slow process which proceeds by leaps and bounds. There is too much noise for annual observations to give a meaningful indication of long-run trends. The Commission’s statement also makes clear that annual measurement is considered desirable.

Why would this be so? In a hypothetical world without uncertainty, the results of aid would be produced fully predictably from inputs via outputs and outcomes to impact. Given such a mechanistic “production process” one could measure results just as well at the beginning as at the end of this process.²³ Clearly, as soon as uncertainty is introduced, early indicators become imperfect predictors of final results. It is then sensible to monitor intermediate results. However, the function of such monitoring is to establish whether the country is still on track. It cannot be a substitute for measuring final results, *i.e.* impact. Indeed, ownership is considered important precisely because there is no unique way to reach particular targets. Basing aid allocation on inputs, outputs or outcomes is therefore fundamentally inconsistent with ownership.

Here donors are caught in a dilemma. They have tried to resolve the dilemma by denying it, as in the Commission’s statement which in fact asserts that the two functions *are* compatible. Adam and Gunning (2002) proposed a distinction between those indicators which are designed to form the basis for the donor’s aid allocation decisions on the one hand, and indicators which are meant to monitor interim progress. This proposal is still relevant.

Trying to combine the two functions, as in the EU practice, is undesirable for two reasons. First, as noted, it undermines ownership. If an intermediate indicator shows a deterioration that calls for an audit in order to establish the reason. If instead the change triggers a reduction in aid, the incentive effect may be perverse, the government being punished when it was on target in terms of ultimate objectives.

This is related to the emerging practice of selectivity. Under selectivity donors active in a particular country are, by definition, in broadly in agreement over the government’s policies as they affect poverty: without such agreement there would be no justification for aid. In addition, there must be an assurance that the government’s general policy stance

²³ Adam and Gunning (2002), p. 2053.

will not change fundamentally in the short run. In these circumstances donors will have to take two, very different decisions: whether the country still qualifies for aid and how much aid it should get.

The first decision requires a judgement as to whether an event has occurred indicating that the government's general policy stance is no longer the one the donors agreed with. Adam and Gunning (2002) described this as "show-stopping" conditions. The important point to note is that performance indicators are *not* designed for that purpose. Measuring achievements in terms of school enrolment, vaccinations or the number of teachers trained makes sense only if the government's policy stance is one which ensures the good policy environment in which donors are willing to support the country. For example, school enrolment in Zimbabwe might be quite satisfactory but donors have decided that the country does not qualify for aid.

As Adam and Gunning stressed for the case of Uganda, one needs to distinguish clearly between indicators which inform the decision to support the country or not and indicators which inform the decision on the scale of that support. Ironically, donors were quite clear on the former, but extremely vague on the latter. It was clear at the time (in 2001) that continued donor support was contingent on the extent of corruption in Uganda and the country's involvement in the war in Zaire. By contrast no "transparent mechanism exists to allow donors to sign clearly the conditions under which the country may expect increased or decreased support (even though such a mechanism may be implicit in the thinking of many donors)."²⁴ The resulting vagueness may be annoying to the government.

The donor position is consistent if one adopts the extreme form of selectivity, i.e. the donor does not believe that he can influence the government by building incentives into the aid contract. However, donors do seem to believe in incentive effects. This makes their position incoherent: the vaguer the link between aid and performance the weaker the incentive offered to the government.

The Commission's second reaction to the objection that its system may penalise a government for an outcome beyond its control is to recognise that "an exogenous shock" may disrupt a government's plans. In this case the indicator concerned is "excluded from the calculation".²⁵ As a practical solution this is impeccable. However, it implies that the insurance function of aid is abandoned.

The EU uses no impact measures. To some extent this is understandable. As has long been recognised, there may be long lags between government actions and, say, poverty reduction. Tying aid to impact in terms of poverty would then reward or punish a government for the actions of its predecessors while giving it little incentive to reduce poverty if it does not expect to be in power when the results of its actions show up as a reduction in poverty. It is therefore understandable that aid contracts do not rely very much on impact indicators. However, in practice they do not use impact indicators at all.

²⁴ Adam and Gunning (2002), p. 2052.

²⁵ European Commission (2005), p. 49.

[Give EU evidence.] This gives distorted incentives. For example, in education a typical indicator is enrolment but there is no attempt to assess what children learn at school. Where in many countries enrolment was expanded very rapidly, quality may have deteriorated. There is shocking evidence on this for Uganda.²⁶ A well-known example is that when the number of bore holes drilled in rural areas was selected as a performance criterion, more effort was expended on drilling bore holes and less on maintenance so that water quality suffered.

This has implication for auditing and evaluation. First, these functions must be independent to reduce manipulation of the performance indicators as much as possible. Secondly, auditing and evaluation should move beyond the agreed performance indicators so as to investigate the relationship between these and final objectives.

6. Impact Evaluation

Aid-supported development activities are evaluated to an unusual extent. Indeed, it is difficult to think of public sector activities on which more information is collected for taxpayers. In most donor agencies, however, evaluations have focused on process. Evaluation reports indicate whether planned activities were undertaken as envisaged, whether objectives (often formulated in vague terms) were reached, whether problems were encountered and, if so, what action was undertaken. There is much to be learnt from such reports but they are silent on what taxpayers really want to know: did the aid work?

Increasingly, formal, econometric techniques are used to address just that question. Such *impact evaluation* tries to establish a convincing counterfactual so that measured results can be compared with what would have happened without the aid-supported activity. Where randomisation is involved (e.g. when a schooling program was implemented gradually and participating schools were chosen randomly) impact evaluation is very similar to the experimental designs used in the sciences. In the absence of randomisation, statistical techniques have to be used to construct the counterfactual. Impact evaluation was first developed in labour economics. Recently it has become very popular in development research.²⁷ In the last few years donors have started experimenting with impact evaluations. In part this reflects their growing awareness that traditional evaluation methods, while certainly useful, do not address the issue of aid effectiveness.

The recent emphasis on budget support, ownership and results-based lending affects this development of aid evaluation in several ways. First, as donors switch from financing projects to budget support they may come to feel that the effectiveness of aid should no longer be assessed at the micro level but at the macro level instead, as in the growth regressions used to derive efficient aid allocations. Secondly, donors might come to feel that if the aid is tied to outcomes there is no need for impact evaluation. Both of these reactions would be mistaken.

²⁶ Adam and Gunning (2002), p. 2050.

²⁷ For an entertaining and fairly non-technical introduction see Ravallion (2001). One of the most famous applications is Miguel and Kremer (2004). See Duflo (2004) for an overview of recent developments.

To start with the last point: if performance indicators record progress, *e.g.* a fall in poverty, aid may be ineffective, for two reasons. First, the fall in poverty may have nothing to do with the aid. Without impact evaluation it would be impossible to avoid falsely attributing the success to aid. Secondly, when the success is in fact correctly attributed to aid, it is possible that the fall in poverty was achieved at very high cost. This points to the need for cost-benefit analysis, which is, of course, quite common in project analysis but which is rare in modern impact evaluations which usually are content to establish whether or not aid has a statistically significant effect. By contrast in the macro literature on aid effectiveness cost-benefit calculations are becoming more common. For example, Collier and Dollar (2002) calculated that on average \$ 1 million of aid had the effect of reducing poverty (as measured by the headcount) by 300 people. Such an average then can be used as a benchmark by which the poverty impact of aid in particular countries can be judged.

A fundamental problem is how aid is to be evaluated if it is given in the form of budget support rather than project finance. The methodology of impact evaluation is suitable only for activities with homogeneous output, *e.g.* an increase in the school enrolment of poor children. It is not designed to evaluate the aggregate of government activities in, say, the health sector. But if donors contribute to a common pool in that sector aid effectiveness obviously requires an assessment of the effectiveness of all activities in the sector. Given the heterogeneity of such activities the sector can obviously not be treated as a single project for the purpose of evaluation.

What is the alternative? In principle, given a reliable description of the activities within a sector it should be possible to draw a representative sample and to apply impact evaluation to each of the activities in the sample. Aggregating the results of the evaluations one would arrive at an estimate of the average effectiveness of aid to this sector. (Note that this approach will not establish the marginal effect; this may well differ substantially from the average effect.) Organising such a sector-wide evaluation is complicated, costly and unattractive for a single donor to undertake. The logic of the common pool approach suggests that such an evaluation is a joint activity of the government and all donors involved in the sector. After all the average effect of a sector activity is the same, irrespective of the source of financing. These issues are beginning to be discussed. As far as I know, no such sector evaluations have yet been undertaken.

7. Conclusion

The proposal for results-based aid was a response to the failure of conditionality. While selectivity in aid allocation does not rely on any incentive effects the question how aid affects the incentives facing the recipient government has recently become prominent, partly as a result of the work of political scientists in the rational choice tradition. In addition, as results-based aid is beginning to be implemented questions about the appropriate design of the aid contract are becoming urgent.

We have noted that in practice the aid contract is incomplete. While many donors have agreed on the performance indicators to be used they have been reluctant to spell out how

much aid would fall or rise with performance. If the contract is to have incentive effects then such vagueness must be avoided. This is an instance of the irony of aid: donors often shy away from the very measures which would make aid effective.²⁸

A key issue in the design of results-based aid contracts is, of course, the setting of targets. If targets are expected to have incentive effects then they must be set at realistic and sufficiently ambitious levels. So far this has been left to consultations between donors and an individual recipient government. We have suggested that this process could be improved by using cross-country regressions to introduce international comparability. The regressions would give the expected level of progress on a particular performance indicator for a country with a particular set of characteristics. The aid contract could then specify how much more the country would receive if this expected level was exceeded or, conversely, how much aid would be cut if performance fell short of this average.

We have suggested that there is a need for independent measurement of performance. This has not yet become a prominent issue precisely because the aid contract has usually been left vague. However, as the relation between performance and aid is beginning to be specified more precisely there are strong incentives to manipulate the data. Independent verification of the performance indicators is therefore becoming increasingly important.

We have noted that there is a case for including partial insurance in the aid contract. The key issue here is that the impact of the shock on performance indicators is usually not known and that it depends in part on actions undertaken by the government, both before and in response to the event. International evidence can be used to reduce uncertainty about the shock's impact. Incentives for the government to mitigate that impact can be built into the aid contract by offering partial insurance rather than full compensation.

The aid contract could therefore be improved in four ways: by indicating precisely how much aid a country will get, depending on performance; by basing target levels on international average performance, controlling for relevant country characteristics; by ensuring independent data collection and auditing of a country's performance; and, finally, by including partial insurance for particular shocks such as droughts.

A corollary is that aid may well be volatile. This would be the case if it included an insurance element. If the aid contract were well designed (in the sense that the government's incentive constraint is satisfied) this volatility would be seen as beneficial rather than detrimental by the government.

Donors have attempted to use performance indicators for two very different functions. In the one case measures indicate whether the country is still on track in the sense that the policy environment is one in which aid can be used productively. In the other case the indicators measure ultimate objectives such as poverty reduction. We have argued that these two functions cannot be combined: donors will need to distinguish between monitoring variables and the performance indicators to which they tie their aid.

²⁸ Gunning (2004).

It is tempting to view results-based lending as a guarantee for aid effectiveness. We have argued that this would be mistaken. There remains a need for impact evaluation (in the econometric sense of the word) to assess whether aid actually makes a difference in terms of the chosen impact indicators. Much of the impact evaluation literature stops at that point. However, there is a need to move beyond establishing impact towards cost-benefit analysis. A much more difficult challenge is to use impact evaluation (designed for analysis at the project level) to evaluate budget support at the sector or national level.

References

- Adam, C.S., G. Chambas, P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning, (2004), 'Performance-based Conditionality: a European Perspective', *World Development*, vol. 33, pp. 1059-1070.
- Adam, C.S. and J.W. Gunning, (2002), 'Redesigning the Aid Contract: Donors' Use of Performance Indicators in Uganda', *World Development*, vol. 30, pp. 2045-2056.
- Adam, C.S., and S. O'Connell (1999), 'Aid, Taxation and Development in Sub-Saharan Africa', *Economics and Politics*, vol. 11, pp. 225-254.
- Azam, J. -P. and J.-J. Laffont (2003), 'Contracting for Aid', *Journal of Development Economics*, vol. 70, pp. 25-58.
- Collier, P. and J. Dehn (2001), 'Aid, Shocks and Growth', World Bank Policy Research Working Paper 2688.
- Collier, P. and D. Dollar (2002), 'Aid Allocation and Poverty Reduction', *European Economic Review*, vol. 46, pp. 1475-1500.
- Collier, P., P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning (1997), 'Redesigning Conditionality', *World Development*, vol. 25, 1997, pp. 1399-1407.
- Collier, P., P. Guillaumont, S. Guillaumont Jeanneney and J.W. Gunning (1999), 'Reforming Stabex', *The World Economy*, vol. 22, pp. 669-82.
- Collier, P. and J.W. Gunning and associates (1999), *Trade Shocks in Developing Countries*, Oxford: Oxford University Press.
- Duflo, E. (2004), 'Evaluating the Impact of Development Aid Program: the Role of Randomized Evaluations', 2nd annual AFD-EUDN Conference, November 25, Paris.
- European Commission (2005), *EC Budget Support: an Innovative Approach to Conditionality*, DG Development, February.
- Guillaumont, P. and L. Chauvet (2001), 'Aid and Performance: a Reassessment', *Journal of Development Studies*, vol. 37, pp. 66-92.
- Gunning, J.W. (2001), 'Rethinking Aid', in: B. Pleskovic and N. Stern (eds.), *Annual World Bank Conference on Development Economics 2000*, Washington, DC: World Bank, pp. 125-144.

- Gunning, J.W. (2004), 'Why Give Aid?', paper presented at the 2nd AFD-EUDN Conference *Development Aid: Why and How?*, Paris, November 25th, 2004.
- Kanbur, R. and M. Tuomala (2003), 'Incentives, Inequality and the Allocation of Aid when Conditionality Doesn't Work: an Optimal Nonlinear Taxation Approach', mimeo, Cornell University, June.
- Miguel, E. and M. Kremer (2004), 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', *Econometrica*, vol. 72, pp. 159-217.
- Ray, D. (1998), *Development Economics*, Princeton, NJ: Princeton University Press.
- Ravallion, M. (2001), 'The Mystery of the Vanishing Benefits: an Introduction to Impact Evaluation', *World Bank Economic Review*, vol. 15, pp. 115-140.
- World Bank (1998), *Assessing Aid: What Works, What Doesn't and Why*, Oxford: Oxford University Press, for the World Bank.