

Comments on "Assessing Multilateral Effectiveness" by Scott

Rick Davies, Thursday, 20 May 2004

Background

1. I came across this paper on the Web, at topics.developmentgateway.org/aideffectiveness/rc/filedownload.do~itemId=398241
The paper includes an invitation to make comments. So here they are.

How will we know when the Multilateral Effectiveness Framework is working?

2. M&E systems get set up for many reasons. They can be seen as rituals, which will only have consequences for those responsible *if they are not developed*. Once developed they don't have much impact. They can also be seen as investment, involving sums of money that could be spent elsewhere. In this case some kind of effects are expected.
3. The paper refers to three kinds of possible effects.
 - On the functioning of the multilateral
 - On funding decisions by DFID
 - On how DFID engages in discussions with bilaterals and other partners on multilateral effectiveness
4. Statements about expected effects on the multilateral focus largely on *proximate* changes, that is, in the relationship between DFID and the multilateral. For example:
 - "enhancing the quality of our engagement as stakeholders or members" (para 2.1)
 - "to provide a basis for dialogue" (para 2.1)
 - "DFID's objective of using MEFF for dialogue with multilateral partnersplaces a premium on a methodology that promotes communication and agreement over shared objectives" (para 10.9)
5. While the focus of the MEFF is on "agency effectiveness" it is noted that these should have a connection with "organisational effectiveness", which refers to the immediate results that can be attributed to an individual organisation directly e.g. schools build, (para 4.3, 4.7). Elsewhere it is noted "The assumption is that effective organisational systems produce consistent standards and behaviour in an organisation, and ultimately, good country level results" (para 4.8)
6. My interim conclusion is that beyond changes in the relationship between DFID and the multilaterals concerned, it will not be easy to see the impact of the MEFF further down the line. However, there are some possibilities, which are discussed below.
7. The paper's position on the implications of the MEFF for DFID funding decisions are ambivalent. It is noted that multilaterals receive almost half of DFID's budget and that funding decisions about multilaterals will require information on the multilaterals effectiveness". However this will be only one of a number of criteria that will be used. While DFID does make specific funding decisions about how much money each multilateral will receive the paper notes "It is not our decision to produce 'league tables' since this could undermine our relationship with the organisations. Taken literally, this means advice originating from MEFF could only have a limited effect on how much funding is received by a multilateral. Any advice that prompted changes in the rank order of amounts of funding received by

multilaterals would require some sense of the rank order (of performance) of the multilaterals, whether expressed in "league tables" or not. This would apply not just to gross funding levels, but also to the relative scale of any annual funding increases or decreases.

8. The MEFF could have a bigger effect on funding decisions, if it was less timid in its ambitions, and was willing and able to make cross agency comparisons. See below.

Managing the relationship between DFID and the multilaterals

9. The paper makes repeated reference to the relationship, and how it needs to be managed. This seems to be taken for granted as a fact of life, essential if anything is to be achieved, in terms of improved performance by the multilaterals.
10. There is a separate section on Effective Partnerships, which is part of the MEFF's tripartite model of domains of agency effectiveness. This section seems to focus on the relationships between the multilateral and a range of other organisations. It does not make any specific reference to relationship with DFID.
11. Given that it is the relationship between DFID and the multilateral, where the MEFF is expected to have its most immediate effects, it would be worthwhile thinking about how to monitor changes in that relationship, in particular. The transparency criteria, from para 8.6 would be one dimension worth attending to. I would also suggest tracking changes in *the ability of the relationship to identify and resolve differences of opinion* (See example in "Does empowerment start at home?" at www.mande.co.uk/docs/empowerment2.htm) This might be a tangible measure of "the quality of our engagement" . Improvements in the ability to identify and resolve differences of opinion could have practical consequences for the scale of impact that DFID's engagement could have on the multilaterals. It would also help prevent stasis and complacency, that a peaceful relationship was necessarily a good relationship.
12. One area where progress might be expected to be made sooner than later would be in the multilaterals ability to cope with comparisons with other agencies (para 2.5), and receiving low cores on rating scales (para 10.6), and to cope with judgements overall (para 10.8)

The MEFF Theory of Change

13. There are number of sections of the paper which outline various assumptions and beliefs. These are theories of how things work, theories of change (ToC). For example, paras 4.5 and 4.6 about the conditions where development aid and humanitarian assistance will have the greatest impact.
14. The MEFF has its own specific ToC, about agency effectiveness: "The *assumption* is that effective organisational systems produce consistent standards and behaviour in an organisation, and ultimately good country-level results. Poor results can usually be traced to weak organisational systems". The MEFF identifies eight organisational systems and three domains of effectiveness for each of these. For each of the resulting 24 performance criteria there are up to four indicators.
15. Another aspect of this theory is the relative importance given to each of the eight different systems and three domains of effectiveness. This is dealt with less

explicitly, but the provisional assumption seems to be that they are all equally important. There is passing reference to the fact that weightings could be given, but no discussion of how they could be assigned.

16. The weakest part of the theory built into the MEFF is the uncertainty about how the results should all add up, how they should be aggregated into meso and macro-level judgements about the performance of a multilateral. There is a reference to rules for aggregating scores, but this is not discussed in this paper. The paper opts for the simplest possible scoring system (3 values) to enable users to perceive "the overall pattern of scores" The priority concern at this stage is the potential effect on the relationship between DFID and the multilateral. This seems to imply that there is no ideal pattern of scores, but that one might be identified through discussion with the specific multilateral. If so, then the Theory of Change is effectively being jointly constructed with the multilateral, on a bespoke basis. That process now needs more elaboration, because there may be some means that can be applied across a number of multilaterals.
17. Getting buy-in to a ToC by the multilateral is obviously important. If they don't agree with it, they are hardly likely to invest significant resources into implementing it or evaluating it's achievement. Getting buy in from the other donors to the same ToC is similarly important. Without it they may be distracting the same multilateral with a different view of where they should be going and what they should be tracking. However, a consensus on the ToC relating to improved agency effectiveness does not necessarily mean that the agreed theory is in the right one. Even when successfully implemented by the multilateral the ToC in the MEFF may have no relationship, or only a marginal relationship to "good in-country results". There needs to be some way of testing it.
18. Trying to do so would be consistent with the emphasis early in the paper about the importance of an MEFF which is *evidence* based (paras 1.2, 2.1 2.4). This was probably meant to refer specifically to the basis of the conclusions drawn about performance in relation to given standards. But to be fair to the multilaterals concerned it should also apply to the performance standards that have been chosen. Are they based on evidence? If they are still assumptions at this stage, will an effort be made to test them? See below.

The multilateral as a machine

19. I was disconcerted to read section 9 on the use of multilateral's own assessment of their effectiveness. Not only about the multilaterals own lack of capacity to assess their effectiveness, but also the conclusion in para 9.3 that these efforts could not be used, and instead DFID would make its own judgements.
20. When we assess the performance of a machine we ask what is it doing and how does that match against what we expect it to be doing. When we assess the performance of an individual or an organisation, we also ask "what does s/he *think* they are doing" A person is expected to have agency; to be aware of choices and to make responsible choices. It is that awareness and responsibility which is the foundation of legal judgements that can make the difference between a death sentence, imprisonment, or freedom. On a more mundane level, it is an individual's (or organisation's) *knowledge* about what has happened which makes the difference between whether what has been done can be changed, avoided in future or replicated. The implication for MEFF is that DFID should be *assessing the multilateral's knowledge* about what it has been doing, and the effects of what it has been doing. That is what matters.

21. Fortunately most organisations know more than the sum total of what has been captured by their M&E systems. Knowledge is also captured in other documents, produced by other sections of the organisation. But more importantly it exists, often in tacit and informal forms, in the heads of people who make decisions about where resources should be allocated.
22. If DFID wants to engage with its multilateral partners, then one means of doing so is by trying to explicate their judgements of their performance, both the criteria they are using, the reasons behind those criteria, and the evidence of achievement on those criteria. This can then be complemented by independent verification by DFID, in the areas of performance that are of the greatest concern. A similar approach was taken with the assessment of a SIDA funded poverty alleviation project in Vietnam (See "A Study Of Perceptions And Responses To Poverty Within The Vietnam-Sweden Mountain Rural Development Programme". at www.mande.co.uk/docs/mrdpppoverty.doc).
23. This does **not** mean abandoning the MEFF structure of 8 by 3 criteria, and the associated indicators. Rather, it provides an orientation as to how they can be used. The following steps could be taken:
- The relative importance of the 8 x 3 criteria can be ranked and explained by the multilateral concerned. Differences with DFID's own views could be identified and may or may not be resolved in the short term
 - The multilateral's relative achievements on each of the criteria can be ranked, explained, and evidenced. Again, differences with DFID's own views could be identified and may or may not be resolved.
 - The MEFF theory (as locally interpreted by the multilateral) would suggest that where priorities are well matched by achievements this should be associated with some positive "country level results" , if the latter is judged in terms of outputs (not longer term effects). And vice versa, less impressive country level results, where the multilateral's achievements in relation to the 8x 3 criteria were weaker. To test this theory the multilateral would need to do some ranking of country level results (region by region being the most practical). As above, this ranking needs to be supported by explicit criteria, explanation, and any available evidence (from M&E systems and/or elsewhere).
24. There is one possible hurdle that needs to be jumped. The multilateral would need to be ready to make some *comparative judgements*. If they are not willing to do so this suggests either very limited organisational capacity, or a relationship with DFID that is lacking in confidence or trust.

Measurement issues

25. The MEFF emphasises the importance of measuring results at the country level. This makes sense for a number of reasons.
- The country is the one common unit of analysis that all multilaterals could be expected to have in common
 - Country governments are one of the most common partners for most multilaterals
 - It is at the country level where there are identifiable constituencies with representative structures than can provide independent feedback on the multilaterals achievements

- Country governments are in most cases also members and financiers of the multilaterals
26. The MEFF proposes the use of ratings, in the simplest possible form. I have suggested rankings rather than ratings. There is a background mini-theory here. The more knowledgeable an organisation is about its priorities and achievements, and the easier it is to use multiple ranks versus three only (which is what the traffic light system can be). If more rank values are used, then more weight can be put on any apparent correlation between two sets of rankings (e.g. priority in relation to achievement). There is another advantage of rankings, which is the poor performance and good performance will be differentiated. With ratings it is possible for all activities of a kind can be given the same rating. That sort of information is effectively useless.
27. The paper makes a point of differentiating between plausible and verifiable claims. This distinction is especially relevant when assessing achievements at the country level (versus internally where evidence is more at hand). However, the emphasis on plausibility risks being seen as a cop out. If it is not being used as an escape clause then there needs to be some means of assessing plausibility. As I have commented on this elsewhere, there are at least two dimensions to plausibility, which can be tracked. One is *detail*. The more detailed a claim of contribution is, the more plausible it is, because to make a detailed claim invites greater risk of disproof. Disproof is easier with a detailed claim compared to a general and unspecific claim. The second is the *public transparency* of the claims being made, for similar reasons. Making public a claim of contribution also involves a greater risk of disproof. Being willing to take that risk suggests confidence in the claim. (See Vulnerability to Disproof, at <http://www.mande.co.uk/archives/200208004.htm>)

Is the MEFF working?

28. Given the discussion above, it is now possible to sketch out in more detail how we might recognise when the MEFF is working. The following might be useful meta-indicators:
- DFID will be able to access progressively more information about the multilateral's performance, from the multilateral itself.
 - There will be improvements in the multilateral's own *knowledge* about its performance. More specifically:
 - There will be a more *differentiated* ranking of its priorities and achievements (in MEFF terms, at least)
 - The scale of real changes in MEFF measures will be in proportion to their priority
 - There will be an increasing correlation between MEFF achievements (weighted by their priority) and country-level results
 - There will be better *evidence* available in support of its claims of achievements, in MEFF terms and at the country level
 - DFID will believe it can accept the multilaterals judgements about its performance (i.e. above) because they are supported by independent verification in the most crucial areas

Comments on these comments are welcomed. Please email rick@mande.co.uk