

**MEASURING AND REPORTING ON VALUE
FOR MONEY**
A conceptual framework for MDBs

Robrecht Renard and Stephen Lister

17 September 2015 (final)



Mokoro Ltd, The Old Music Hall, 106-108
Cowley Road, Oxford, OX4 1JE, UK
Tel: +44 (0) 1865 403179
Email: mokoro@mokoro.co.uk
Website: www.mokoro.co.uk

Foreword

This paper was commissioned by the African Development Bank (AfDB) on behalf of a group of multilateral development banks that have been considering how to strengthen their approaches to assessing and reporting on value for money. In addition to the AfDB, the group includes the Asian Development Bank, the Inter-American Development Bank, the European Bank for Reconstruction and Development (EBRD), the International Fund for Agricultural Development, and the World Bank.

An earlier version of this paper was discussed at a seminar in London, hosted by the EBRD, on 29-30 June 2015.

The final paper is being published as a valuable contribution to an important debate, but the views expressed remain those of the authors and should not be attributed to any of the sponsoring agencies.

The authors would like to thank the MDB staff who provided background documents and gave useful advice by telephone and at the June 2015 workshop. Special thanks to Simon Mizrahi for shepherding the process, and to Stephanie Allan for excellent support. Responsibility for any remaining errors, and for the opinions expressed, remains with the authors.

CONTENTS

1. Introduction and summary	1
2. A conceptual framework	4
3. Project-level VFM tools	11
4. Applying VFM at aggregate level within and across MDBs	26
5. Recommendations	34
References	37
Acronyms and abbreviations	40

BOXES

Box 1	The DAC Evaluation Criteria	6
Box 2	Correspondence table of efficiency concepts	10
Box 3	A classification of efficiency tools for the health sector	12
Box 4	A classification of efficiency tools	18
Box 5	Changing perspectives on efficiency tools	23
Box 6	MDBs’ Common Performance Assessment System	31
Box 7	A comprehensive VFM assessment matrix	33

FIGURES

Figure 1	Alternative aggregations for VFM analysis	2
Figure 2	Efficiency and effectiveness in the economics terminology	8
Figure 3	VFM and the logical chain in DFID terminology	9

1. Introduction and summary

Background

1. The multilateral development banks (MDBs) and their shareholders share a concern that the institutions be effective (achieve their objectives) and efficient (make good use of resources in doing so). This is a matter not only of demonstrating that they are good at what they do, but also of enabling them to learn from experience and improve over time. The different MDBs have different mandates and work in different contexts, but there are enough similarities to suggest that they could fruitfully use similar analytical techniques and compare aspects of their performance.

2. Recent discussions amongst the MDBs have been stimulated by suggestions that they might develop common “value for money” (VFM) metrics. *Value for money* is an intuitively appealing term that embraces both the effectiveness and efficiency dimensions of performance. However, it is susceptible to diverse and partial interpretations that may not lead to the best decisions about the allocation and use of resources. It can be helpful to have simple rules of thumb, but it is important also to understand their limitations and avoid making flawed decisions based on flimsy analysis.

Objectives and structure of this paper

Approach

3. Delivering VFM is central to the MDBs’ mandates. To develop an improved understanding of VFM, five multilateral development banks—the African Development Bank, Asian Development Bank, Inter-American Development Bank, European Bank for Reconstruction and Development (EBRD), and World Bank—and the International Fund for Agricultural Development agreed to jointly commission this concept paper to inform a discussion on assessing and reporting on VFM.

4. MDBs are complex organisations with complex portfolios. They also have separate mandates that differentiate their work. Thus four of the agencies involved in this discussion have geographic mandates, while one has a sector mandate; most are oriented towards the reduction of poverty, but the EBRD’s objectives are framed in terms of furthering progress towards market-oriented economies. It is therefore highly unlikely that any one simple metric will fit them all.

5. This paper aims to assist the debate about MDB VFM by clarifying the concepts involved and suggesting how a common approach may be developed. Such an approach will need to take account of the different contexts in which efficiency/effectiveness analysis may be applied, and of the different types of decisions for which a VFM perspective may be sought.

6. VFM analysis may serve three complementary purposes: (a) to support accountability (of MDBs to their boards and to wider stakeholders); (b) to support learning, to strengthen future use of resources; and (c) to support managerial decision-making, again with the aim of optimising MDB performance. These purposes imply different time-scales: learning from ex-post evaluations may be a long-term endeavour; formal accountability has regular cycles (e.g., annual reports and periodic replenishments); managerial decision-making has a day-

to-day dimension, and decisions frequently have to be made not with ideal information but with the information that is actually available.

7. The idea of comparing VFM across MDBs implies many levels of aggregation. Figure 1 illustrates some of the possibilities within a single MDB:

A: The obvious unit for VFM analysis is the individual project, following the causal chain through from inputs to final outcomes.

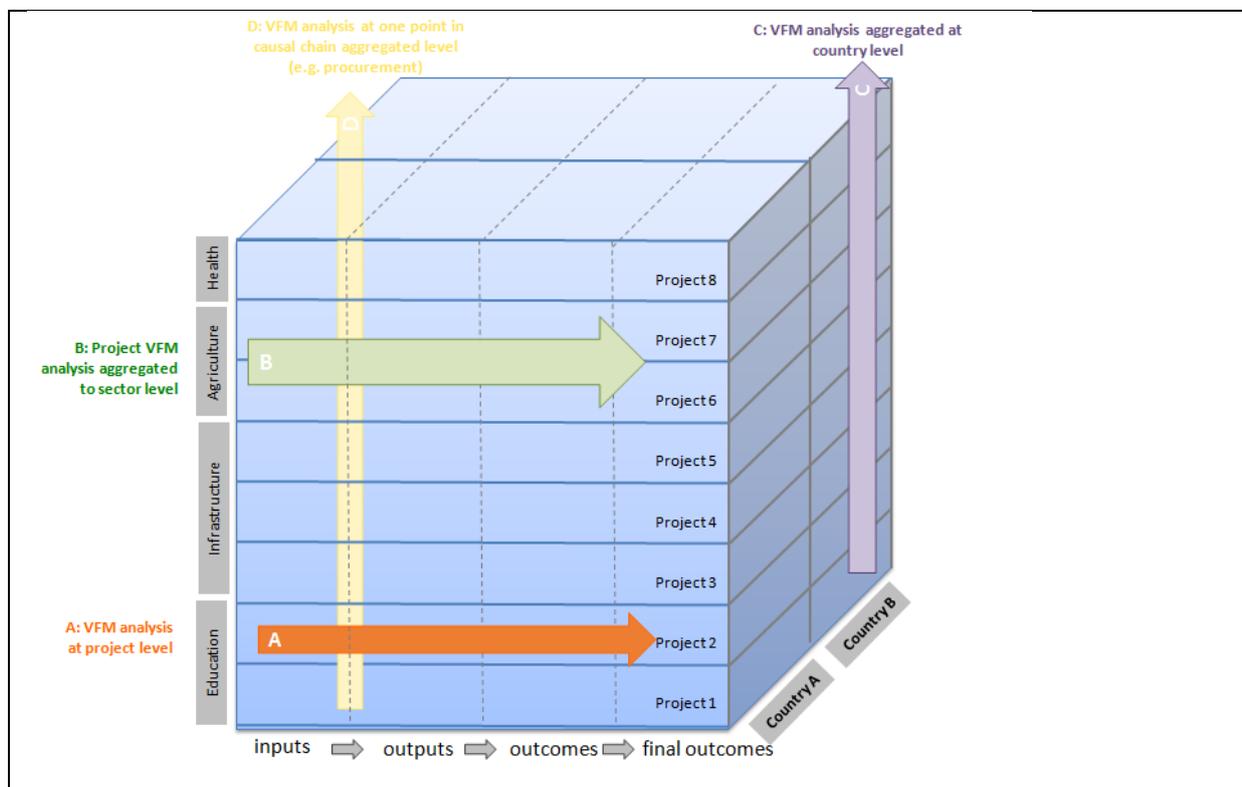
B: It is sometimes feasible to take such analysis up to the (sub)sector level, but making direct comparisons between projects in dissimilar sectors is far more difficult.

C: Assessing VFM at the country portfolio level would require aggregation of dissimilar sectors.

D: At the same time, an MDB may wish to conduct VFM analysis across all its activities at some of the lower levels of the causal chain (e.g., considering the VFM of its procurement or personnel management systems, or considering the efficiency of its project design and approval processes).

8. As the paper proceeds, we consider the practical possibilities for different levels of analysis and aggregation both within and across MDBs, and for combining or comparing information from different VFM perspectives.

Figure 1 Alternative aggregations for VFM analysis



Conceptual framework

9. In Section 2 we present a conceptual framework for assessing VFM. VFM is a familiar term from the literature on public management, but if it is to be applied to the field of development finance the concept must be made consistent with the vocabulary and analytical tools used there. The UK Department for International Development (DFID), which is an important advocate of VFM analysis, proposes its own framework of the “3Es”—“economy, efficiency and effectiveness”—to bridge the two vocabularies, but we demonstrate that this is not very helpful. We propose an alternative framework that is more firmly grounded in economic theory and more in line with the evaluation vocabulary used by the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD). This alternative framework does not change the underlying arguments developed by DFID and other advocates of VFM, but it may help to overcome some of the confusion and hesitation that surround it. The section draws two conclusions: (a) that VFM is best understood in terms of efficiency as defined in welfare economics, and (b) that we must first establish effectiveness (what has been achieved by the intervention?) before we can address efficiency (was it worth the effort?).

10. These two conclusions have important implications. This notion of VFM/efficiency can be established at different levels in the logical framework of an intervention. VFM can only be as convincing as the underlying effectiveness analysis. When we move from activities to output to intermediate and then to final outcomes, the money or cost side of VFM becomes more complex to measure, as for instance costs borne by participants have to be included. On the value side, the difficulty of measuring results increases even more, often dramatically so, as we move up the logical chain. There is therefore a danger that a VFM approach that ostensibly aims at the whole logical chain will focus in practice on the lower elements, resulting in cost-cutting at activity and output levels in ways that may or may not be consistent with an optimal pattern of expenditure.

Understanding VFM at project level

11. In Section 3 we discuss project-level VFM techniques. Although (as we will explore below) the VFM of a portfolio is not just the sum of the VFM of its component interventions, an overall VFM assessment of an MDB portfolio should consider these project-level components.

12. We argue that ideally project-level tools should satisfy the “Drummond criteria”: that is: (a) both the consequences of an intervention (the “value” in *value for money*) and the costs (the “money” in *value for money*) must be examined, and (b) two or more alternatives must be compared. The section demonstrates that the tools best suited to the assessment of VFM are rooted in economics: cost-benefit analysis, cost-effectiveness analysis, and cost-utility analysis. These techniques offer the most realistic prospects of comparing VFM across projects, although the difficulties increase with the dissimilarity of the projects to be compared.

13. The section also recalls the rise and fall of cost-benefit analysis at the MDBs from the 1970s onwards, and suggests what lessons can be drawn from this experience. The section concludes that rigorous efficiency tools can be used to great effect at project level. They can also be aggregated, but only convincingly so up to the subsector level (fighting malnutrition, primary education, disaster prevention, etc.). The good news is that such subsector

aggregation allows important lessons to be drawn, and also constitutes an element of accountability for the MDBs involved. But pushing such aggregation up to the level of a whole MDB portfolio is not realistically possible. For this reason, MDBs’ VFM assessments will need to go beyond aggregated project-level results.

VFM at portfolio level

14. Section 4 takes the discussion fully to the portfolio level. There is a need to develop appropriate additional indicators, not only because of the near-impossibility of aggregating convincingly from the project level upwards, as argued in Section 3, but also because there are some inherent weaknesses in project-level indicators from a portfolio perspective. This section argues that cost-benefit analysis and similar microeconomic approaches cannot capture the following aspects:

- (a) Fungibility: if “what you see is not what you get,” then evaluating “what you see” is rather futile, at least for accountability purposes.
- (b) Systemic effects related to the unintended political effects of donor funding (especially in aid-dependent countries with poor governance); the possible detrimental effects arising from currency overvaluation; and the undermining effect of poorly aligned and harmonised donor interventions on national administrative capacity (which is at the heart of the 2005 Paris Declaration).
- (c) Donor portfolio considerations, in particular regarding the different role of development finance in fragile, poorly governed countries.

15. All of this lends weight to the conclusion that, while MDBs need to insist on rigorous economic analysis as the benchmark for VFM assessments at the project level, such analysis needs to be linked to a broader set of tools to assess whether MDBs are reaching their strategic and managerial goals.

Recommendations

16. Section 5 draws together a set of recommendations for how MDBs can approach assessment of VFM holistically – from the perspectives of accountability, learning, and managerial decision-making. The recommendations draw from all the preceding sections and also highlight the need to make any VFM metrics part of internal decision-making and learning over time, building, to the extent possible, on existing systems for comparing MDB performance.

2. A conceptual framework

Introduction

17. MDBs need a common and adequately rigorous vocabulary for addressing VFM. Unfortunately, the term *value for money* is often used loosely, and some proposed terminologies have introduced more confusion than clarity, and more ambiguity than rigour. The term has been borrowed from the field of public management, where it is used without being given a precise technical definition (see, for example, OECD 2015). But development assistance is a significantly different field:

- (a) Development institutions’ portfolios can be mostly broken down into projects and programmes whose societal value can in principle be traced. In contrast, public administration spending largely consists of salaries and working costs in support of a variety of public services whose societal value is much more difficult to establish.
- (b) However, the *raison d’être* of public administrations is not questioned, and the emphasis is more on what is under management’s control. In contrast, development assistance is contested, and VFM analysis needs to address the question of whether the assistance is actually helping the poor, as well as whether it does so at reasonable cost.

18. VFM is a relatively new addition to the vocabulary in the development community. For instance, it does not appear among the DAC evaluation criteria of relevance, effectiveness, efficiency, impact, and sustainability. Nor is it mentioned in the “Glossary of Key Terms in Evaluation and Results Based Management” (DAC 2002).

19. We therefore propose terminology that:

- (a) Adheres as much as possible to the standard DAC terminology used in evaluations. The DAC evaluation criteria have wide currency, and it is desirable for *ex ante* appraisal and ongoing monitoring to use the same concepts as *ex post* evaluation. Similarly, the logical hierarchy from inputs through outputs to outcomes and impact is a standard tool in both project preparation and evaluation, and is also invaluable in interpreting various relationships between cost and value. Unfortunately, two key terms in the DAC lexicon have acquired ambiguous meanings; as discussed below, rigour requires that these ambiguities be resolved.
- (b) Is consistent with the way efficiency is used in economics, since economics is the discipline best able to bring rigour to VFM calculations.

20. We demonstrate that VFM can best be interpreted in terms of the relationship between cost and effectiveness (i.e., efficiency) at each level of the logical framework. However, we also note that it is, in practice, much easier to calculate efficiency at lower levels of the logical framework than at outcome levels, and this can easily introduce a bias towards cost-cutting rather than value-maximisation in the name of “value for money”.

The DAC evaluation criteria

The basic criteria

21. We reproduce a familiar presentation of the DAC criteria below (see Box 1) and then draw attention to two areas of potential ambiguity: the definition of *efficiency*, and alternative usages of *impact* and *impact evaluation*.

Box 1 The DAC Evaluation Criteria

Criterion	Definition and Sample Questions
Relevance	<p>The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor.</p> <p>Useful questions:</p> <ul style="list-style-type: none"> • To what extent are the objectives of the programme still valid? • Are the activities and outputs of the programme consistent with the overall goal and the attainment of its objectives? • Are the activities and outputs of the programme consistent with the intended impacts and effects?
Effectiveness	<p>A measure of the extent to which an aid activity attains its objectives.</p> <p>Useful questions:</p> <ul style="list-style-type: none"> • To what extent were the objectives achieved / are likely to be achieved? • What were the major factors influencing the achievement or non-achievement of the objectives?
Efficiency	<p>Efficiency measures the outputs – qualitative and quantitative – in relation to the inputs. It is an economic term which signifies that the aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted.</p> <p>Useful questions:</p> <ul style="list-style-type: none"> • Were activities cost-efficient? • Were objectives achieved on time? • Was the programme or project implemented in the most efficient way compared to alternatives?
Impact	<p>The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions.</p> <p>Useful questions:</p> <ul style="list-style-type: none"> • What has happened as a result of the programme or project? • What real difference has the activity made to the beneficiaries? • How many people have been affected?
Sustainability	<p>Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.</p> <p>Useful questions:</p> <ul style="list-style-type: none"> • To what extent did the benefits of a programme or project continue after donor funding ceased? • What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?
<p>Source: http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm</p>	

Ambiguities around impact

22. The definition of *impact* in Box 1 is consistent with the hierarchy of standard logical frameworks, which proceed from activities/inputs, through outputs and outcome to final impact. Unfortunately, *impact* is also regularly used in two other senses: (a) as a synonym for results (or effects) that can appear at any level in the logical framework; and (b) in the

context of advocacy for (rigorous) *impact evaluation*, when this term is meant to distinguish evaluations that (e.g., through randomised control trials) judge results against a rigorously defined counterfactual from evaluations that use less rigorous techniques. Confusingly, impact evaluations in this sense can be used to establish results at output and outcome levels, not just at final impact level. The ramifications of these ambiguities on the term *impact* term are well described in an annex to the Evaluation Cooperation Group’s “Big Book” (ECG n.d.).

23. It makes sense, therefore, to be frugal with the term *impact* (often *results* or *effects* is the intended meaning), and to be very explicit when the technical sense of *rigorous impact evaluation* is intended.

Competing definitions of efficiency

24. The DAC uses two competing notions of efficiency. In Box 1 above *efficiency* is defined as a relation between inputs and outputs. This is a more restrictive definition than the one used in economics, where inputs can be related to any level of results, and preferably to the higher ones in the logic chain. Although the efficiency definition in Box 1 is familiar, it is not the definition of efficiency that appears in the DAC “Glossary of Key Terms in Evaluation and Results Based Management”: “Efficiency is a measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results” (DAC 2002). In line with the logical framework terminology, *results* in this definition can be activities, outputs, outcome or impact.

25. The definition in Box 1 acknowledges that *efficiency* is a term derived from economics, but implies that efficiency analysis is more narrowly focused on the relationship between inputs and outputs. This ambiguity is pervasive, as many donors seek to apply more restrictive terminology (as in the 3Es terminology of “economy, efficiency and effectiveness”, which we discuss below).

26. The broad definition should be preferred because it allows a consistent application of efficiency techniques to answer two key questions with which MDBs are properly concerned: (a) “are we doing things right?” (operational efficiency of development agencies) and (b) “are we doing the right things?” (overall allocative efficiency of interventions, which includes but goes beyond operational efficiency). At the same time, the broad definition helps to explain the relationship between the various partial definitions of efficiency, as well as the significance of VFM.¹ We develop these points in the following paragraphs, which explore the relationship between efficiency and effectiveness and then contrast this approach with the 3Es terminology.

The relationship between efficiency and effectiveness

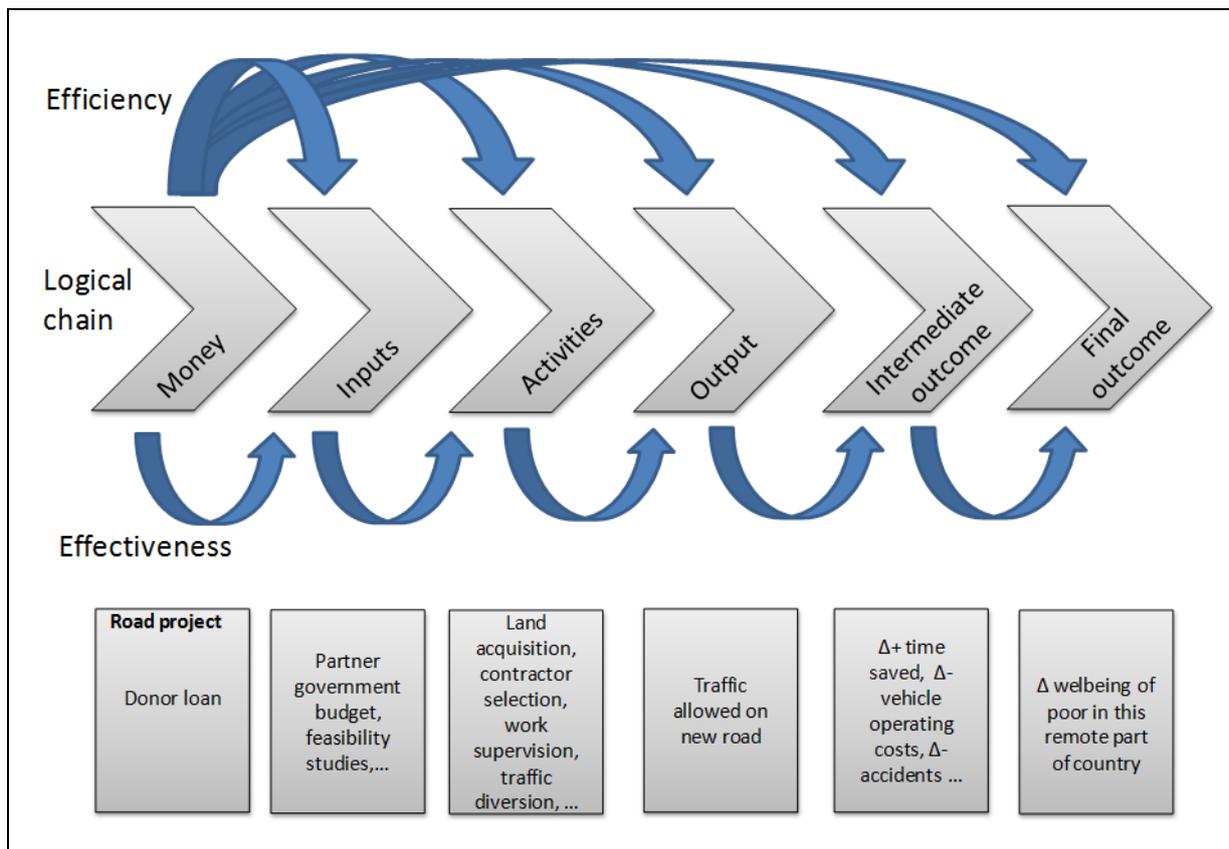
27. Figure 2 below presents *effectiveness* (achieving the goals that were set out) and *efficiency* (achieving these goals at lowest possible cost, also compared to alternatives) as economists understand these terms. We differentiate along the logic chain of an intervention, a notion that is widely used in the development community. We use an

¹ The broader definition is also endorsed by most DAC evaluation experts interviewed for a BMZ study on efficiency (Palenberg 2011:10).

example of a road project financed by an MDB to illustrate how the concepts of effectiveness and efficiency can be distinguished at different stages of the logic chain.

28. In Figure 2, the lower part of the diagram shows five levels of effectiveness for a typical road project. For instance, at the input level, the completion of a technical and financial feasibility study is one sign of effectiveness. Similarly, the selection of a private contractor for the road construction is an indicator that at least one of the activities has been effectively realised. At the output level, an indicator of effectiveness is whether car traffic is allowed on the newly completed road. At the intermediate outcome level, effectiveness is gauged by verifying whether drivers use the new road and in so doing save time on the journey and reduce their vehicle operating costs. And at the final outcome level, given the fact that the MDB financing this project targets the poor in particular, effectiveness may focus on the question of whether the existence of the road increases poor people’s incomes through better market integration, employment creation in a previously isolated area, and so forth.

Figure 2 Efficiency and effectiveness in the economics terminology



29. It is clear from this example that lower (earlier) stages of effectiveness are by themselves not sufficient to justify the project. A road project may be completed and intensively used (output), but the expectation that the number of road accidents per km travelled will decrease (intermediate outcome) may not materialise because of faulty design or lack of safety measures. If in addition economic expansion in the previously isolated area does not take place because of lack of private investment, the poor may not be better off with the project than without the project (final outcome). This suggests a natural hierarchy

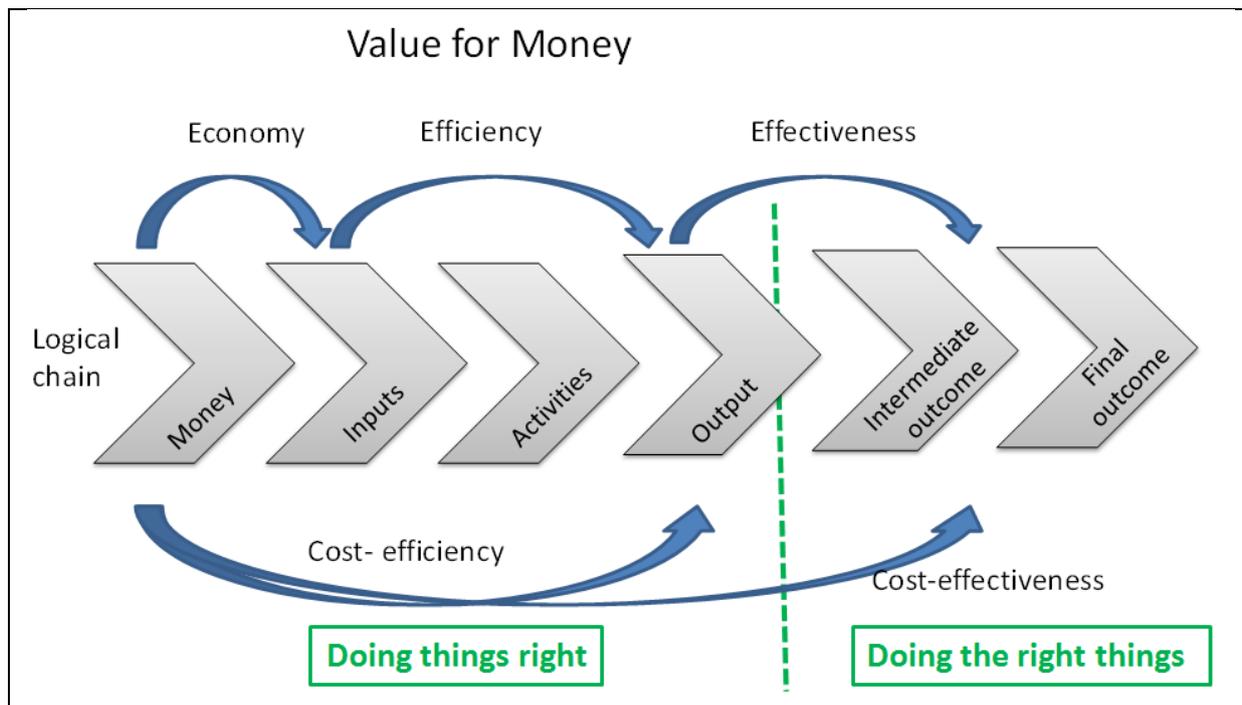
between the different levels of effectiveness: lower-stage effectiveness indicators are a necessary but insufficient condition for next-stage effectiveness. Effectiveness at input, activity or output level is thus useful to check whether an intervention has a chance of succeeding where it really matters, that is, at outcome levels, but does not show that it is sure to do so.

30. Figure 2 also depicts five corresponding levels of efficiency. At input level, the question is not only whether a feasibility study of good quality has been produced, but also whether it has been acquired at a reasonable cost, compared with suitable benchmark costs for similar studies in comparable circumstances. Similarly, at activity level, the cost of supervision is justified by the quality of construction it helps guarantee. At output level, the cost per km of road should not exceed that for similar roads in comparable circumstances. At intermediate outcome level, the benefits of the road in terms of time saved, reduced vehicle operating costs and reduced accidents should be greater than the costs. And at final outcome level, from the perspective of the poor, the economic advantages of the new road project should exceed the costs and compare well with alternative means of improving their lot. It is clear that lower-level efficiency is neither a necessary nor a sufficient condition for higher-level efficiency. For instance, the fact that a project is efficient at intermediate outcome level (e.g., a road succeeds in reducing vehicle operating costs) does not guarantee that it is also efficient at final outcome level (its costs may not be justified in terms of the difference it makes to the well-being of the poor).

Efficiency and the 3Es terminology

31. In the donor documents relating to VFM, the restrictive definition of *efficiency* seems to dominate, often in the 3Es version of economy-efficiency-effectiveness that is promoted by DFID but is not used by MDBs (see Figure 3 below).

Figure 3 VFM and the logical chain in DFID terminology



32. The contrast between Figure 3 and Figure 2 is striking. Efficiency in Figure 3 is restricted to the output level. At input level it is replaced by economy, which may be defined as “reducing the cost of resources used for an activity, with regard to maintaining quality” (Jackson 2012). Economists will not have a problem with this additional term, which is related to cost minimisation, a subcategory of efficiency that is appropriate when two or more alternatives produce exactly the same effects. More surprising and confusing is that at levels above the output level the 3Es terminology switches to effectiveness. From the perspective of the definitions used in economics (Figure 2), this conflates *effectiveness* (the effect of an intervention) and *efficiency* (whether that effect is worth the cost involved in obtaining it), and it restricts efficiency to the output level, thereby robbing it of its most important applications (see also Renard 2014).

33. Box 2 below compares the contrasting terminologies of Figure 2 and Figure 3, demonstrating that all the 3E terms can be directly translated into the more consistent and straightforward terminology of economics. The underlying argument in favour of VFM is not affected by the switch, but the calibration with the existing DAC and economics vocabulary is improved, and the unnecessary confusion introduced by the 3Es is dispelled.

Box 2 Correspondence of efficiency concepts

Economics terminology (Figure 2)	3Es terminology (Figure 3)
Efficiency at input stage	Economy
Efficiency at output stage	Cost-efficiency
Efficiency at intermediary or final outcome stages	Cost-effectiveness
Effectiveness at output stage	Efficiency
Effectiveness at intermediary or final outcome stages	Effectiveness
All of the above	Value for money

34. This mapping shows that the broad definition is actually rather simple. There are additional reasons to prefer it:

- (a) It is consistent with the economics literature on efficiency (as we elaborate in Section 3 below).
- (b) Conversely, the definition that strictly limits *efficiency* to a relationship between inputs and outputs suggests that the relationships of inputs to intermediate outcomes and of inputs to final outcomes² are not the realm of efficiency analysis. The narrow definition fails to acknowledge the simple but important fact that the notion of efficiency applies to all stages of the logical framework beyond input, rather than just to output, and that there is some inherent hierarchy between these stages, with efficiency assessment at higher stages being the most meaningful.

² Of inputs to impact in the DAC terminology.

- (c) Furthermore, the narrow definition obfuscates the distinction in evaluation theory between two questions: what is the effect of the intervention at a certain stage in the logical framework (effectiveness), and could we have achieved the results at lower cost by choosing another alternative (efficiency)?

Conclusions

35. VFM needs to be understood through the underlying economic concept of efficiency, which concerns the relationship between costs (not exclusively financial) and the results of an intervention (which can be monitored at every level of the causal chain up to final impact). At every level, efficiency can be calculated only if both cost and results data are available. In other words, efficiency calculations always require information about effectiveness as well as cost. Frequently, and especially at intermediate and final outcome levels, data on effectiveness are even more difficult to acquire than adequately disaggregated data on costs. Efficiency calculations at lower levels of the results chain are easier to do, and can be very useful for programme management. However, they need to be treated with care, because efficiency at lower levels is not always consistent with efficiency at subsequent stages. (For example, less expensive raw material may result in a lower quality, and less effective, product.)

36. Moreover, attention to the broad definition makes it clear that “*value for money*” is properly understood as efficiency at the final outcome level. We note that the UK Independent Commission for Aid Impact (ICAI) reached a similar conclusion:

ICAI’s view is that effectiveness and value for money are inextricably linked: how can a programme be value for money if it is not effective; and if there is poor value for money, is the programme being as effective as it could be?

ICAI recognises that aid organisations, including DFID, take a range of different approaches to defining and ensuring effectiveness and value for money. The key aim of our reviews, however, will be to assess the results that have been achieved on the ground. In our view:

- effectiveness involves achieving a sustained impact for intended beneficiaries; and
- value for money is the best use of resources to deliver the desired impact (ICAI 2011, ¶2.17-2.18).

3. Project-level VFM tools

Introduction

40. This section builds on the concepts and definitions proposed in Section 2. It focuses on intervention-level efficiency analysis because projects/programmes/operations are MDBs’ standard unit of analysis and management. From the discussion in Section 2 it is clear that “*value for money*” is properly understood as efficiency at the final outcome level, and we will continue to use the term *VFM* in this sense. It will, however, become clear that if the ambition is to assess whole portfolios or, even more ambitiously, to rank development agencies on their overall VFM, it will be necessary to broaden the toolbox.

41. We take it for granted that the MDBs wish to measure VFM as accurately and objectively as possible, so as to make a convincing case to their boards, to their donors, and

to public opinion, relying where possible on quantitative (or at least transparently objective) techniques. Among other considerations, we need methodologically robust and mostly quantitative tools to be able to aggregate across projects and perform meta-evaluation.

42. Successive subsections consider (a) the choice of tools for measuring VFM; (b) the practicality of efficiency analysis; (c) the potential and limitations of aggregating VFM metrics of individual interventions; (d) ways the accountability perspective of VFM may clash with the learning objective; and (e) the costs and benefits of performing VFM analysis.

How to measure VFM

Criteria for selecting efficiency tools

43. In a well-known handbook on the economic evaluation of health projects, Drummond et al 1997 classify instruments for efficiency analysis. In line with their analysis, we suggest that rigorous VFM tools for individual interventions must satisfy two conditions: (a) both the consequences of an intervention (the “value” in *value for money*) and the costs (the “money” in *value for money*) must be examined, and (b) two or more alternatives must be compared. With respect to this second condition, the counterfactual is one of the alternatives; an intervention must at least be compared with what would have happened in its absence. The results are presented in Box 3 below.

Box 3 A classification of efficiency tools for the health sector

Both costs and consequences of the alternatives examined?				
		NO		YES
		Only consequences (value)	Only costs (money)	
Comparison of two or more alternatives?	NO	Outcome description	Cost description	Cost-outcome description
	YES	Effectiveness analysis	Cost analysis	Full economic evaluation: <ul style="list-style-type: none"> - Cost-benefit analysis - Cost-utility analysis - Cost-effectiveness analysis - Cost minimisation

Source: adapted from Drummond et al 1997, table 2.1 p10.

44. Applying these twin conditions leads to the four tools listed in the bottom right corner of Box 3. If the MDBs accept these “Drummond conditions” for assessing at least a core selection of their individual interventions, as we suggest they do, they should start looking there.

Cost-benefit analysis

45. Cost-benefit analysis (CBA) consists of a systematic comparison of the disadvantages (or costs, or money) and advantages (or benefits, or value) of an intervention. For the sake of both scientific rigour and precision, costs and benefits are expressed in commensurate quantitative terms. Costs and many benefits (e.g., fuel savings and decreased cost of vehicle maintenance in road projects) are naturally expressed in money terms. Other benefits can be expressed in monetary terms without too much controversy. Time saved by car users, for instance, can be estimated by multiplying the hours saved by an average hourly wage, to express the production that can be gained by not being stuck in traffic.

46. However, road projects also save lives, and here monetisation is more controversial. Three techniques can be used in such cases. The *human capital approach* calculates the income a person would have earned in the additional life years, and postulates this as a minimum estimate of the value of this life. The *stated preference approach* uses interview techniques to let people indicate the monetary values they attach to a set of hypothetical situations involving life, illness and death. The *revealed preference approach* studies the actual behaviour of individuals in risk-related markets. The revealed preference approach can also be applied to governments or international donors: both spend (or force individuals to spend) money to reduce the risk of accidents and death. Safety standards in new cars and buildings, technical controls of used cars, stiff penalties for traffic rule violations, and spending on hospital ambulance services are all examples. Implicitly, by spending so much, and not more, on any such measure, public authorities (and those who fund them) assign a monetary value to human life that can in principle be calculated and applied to the new road project.

47. Does CBA respect the Drummond conditions? The first condition is not in question, as both the costs and the benefits are studied. What about the second condition? CBA is sometimes used to compare projects, such as alternative road designs, or the societal profitability of different types of education, and then two or more alternatives are explicitly compared. When CBA is limited to the study of a single intervention, however, there may be an impression that no counterfactual is being considered, but this is not the case. On the cost side, the basic notion used in CBA is that of opportunity cost. Using certain resources, such as land, labour, or electricity, means that these same resources cannot be used elsewhere, and this is what needs to be expressed on the cost side: the forgone value of what these resources would have delivered in the absence of the project. The market price or the cost to the project agency may or may not be a good expression of this. For instance, if labour markets are distorted, the opportunity cost of (un)skilled labour may be higher or lower than the wage actually paid. On the benefit side, the methodology of CBA is similarly geared to identify the counterfactual. A simple example is an import-substituting project that produces malaria drugs. At first sight one could believe that the benefit of such a project is the value of the drugs, as expressed by consumers' willingness to pay, or, maybe more accurately, as the benefits to society of the longer and healthier lives of those taking the drugs. But if, in the absence of the project, identical malaria drugs would have been imported, the counterfactual is different: the benefit of the project is the foreign exchange saved by no longer having to import malaria drugs. And these are precisely the calculations that textbooks on CBA prescribe.

48. Several formal decision rules are available in CBA. The simplest is just to subtract the costs from the benefits. A project passes the efficiency test if the resulting net benefit is positive. Because projects often have a time horizon of several years, discounting is applied to express all costs and benefits in their equivalent value at the start of the project. This leads to the net present value (NPV) criterion: if the NPV is positive, the project passes the efficiency test. Alternatively, the (discounted) benefits can be divided through by the (discounted) costs to obtain the (discounted) benefit-cost ratio (BCR). If the BCR exceeds unity, the project passes the efficiency test.³ Finally, it is possible to do a sensitivity analysis by calculating the switching value for one of the economic parameters used in the calculation. The best known such criterion is the internal rate of return. If the internal rate of return exceeds the chosen discount rate, which expresses the opportunity cost of capital invested in the project,⁴ the project passes the efficiency test. Another example is the domestic resource cost, sometimes used in the evaluation of agriculture projects in developing countries. A project passes the efficiency test if the domestic resource cost is lower than the (shadow) exchange rate. All these are simple algebraic transformations of the NPV rule, and all will normally give the same yes-or-no answer to the question of whether a project provides VFM.

49. Sometimes a simple yes-or-no answer is insufficient. The decision on several projects may be linked. They may be mutually exclusive (as when alternative road options are compared: by choosing one option, other options are ipso facto excluded) or subject to a constraint (as when an MDB has identified a series of not mutually exclusive projects in different parts of the country, all with positive NPV, but exceeding the total investment budget that it has set aside). Among mutually exclusive projects, the rule is to choose the one with the highest NPV. In the case of a single-input constraint, the rule is to calculate the ratio of the NPV to the number of the constrained inputs required, rank projects in decreasing order, and select projects until the input is exhausted.⁵ Here is a crucial link with cost-effectiveness ratios (CERs) discussed below: it can be shown that under certain not-too-restrictive assumptions a CER gives the same ranking as the NPV per unit of the constrained input.

Cost-effectiveness analysis (CEA)

50. When the benefits of a project cannot be expressed in monetary terms, but it is possible to express them in some other metric (e.g., number of children educated, kilometres of road constructed, number of people attending health clinics, number of lives saved), then a CER can be calculated, expressing the cost per unit of result. Projects with lower CERs or unit costs are to be preferred to projects with higher CERs. CERs of different projects can be meaningfully compared as long as the metrics used in the numerator (e.g., discounted 2015

³ The inverse of the BCR, the cost-benefit ratio (CBR) can also be calculated; a project passes the efficiency test if the CBR is less than unity.

⁴ In fact, the discount rate is often called upon to express other things as well: inflation, risk aversion, and intertemporal equity. Obviously a single parameter cannot satisfactorily express all of these. It is beyond the scope of this paper to discuss the best solution to this conundrum, but the reference illustrates how controversial the selection of discount rates can be.

⁵ If there is more than one input constraint (say, both a budget constraint and a lack of critical MDB staff to implement all desirable projects), a more complex mathematical algorithm is needed to solve the allocation problem.

US dollars) and in the denominator (e.g., kilometres of road constructed) are the same in all the cases compared. We have seen above that ratios are also used in allocative efficiency analysis in the form of BCRs or their reverse (CBRs). CERs are closely related to CBRs. The difference between the two is that in the CBR denominator benefits are expressed in money terms, whereas in the CER denominator they are expressed in a nonmonetary metric. A CER is therefore simpler to calculate and may also be less controversial, because no monetisation is involved.

51. The CER is closely related to the CBR but has three drawbacks:

- (a) It lacks the CBR's ability to establish the desirability of an intervention. To say that a project is worth doing because the cost expressed in dollars (CBR numerator) is lower than the benefits expressed in the same dollars (CBR denominator) makes intuitive sense. But it makes no sense to say that it is worth doing because the cost expressed in dollars (CER numerator) is lower than the benefits expressed in educational achievements (CER denominator). In this way, in switching from CBR to CER we lose the capacity to perform what is sometimes labelled *allocative efficiency* analysis, although CEA allows us to assess *technical* efficiency (Boardman et al 2011, p484). This does not matter if we are convinced that the competing interventions we are studying all provide VFM in their own right, and that no calculation is needed to justify them individually; in that case the only question is how best to rank them. An illustration might be health projects that all have a cost per human life saved that is less than \$5,000. Presumably most governments, MDBs and donors will accept that this is excellent VFM.
- (b) The CER, like the CBR, is an appropriate ranking device in case of a budget constraint. When there is no budget constraint but projects are mutually exclusive, the CER ranking is only correct if all projects are scalable so as to provide the same level of results, such as employment created.
- (c) It must be possible to express all significant benefits either in monetary terms (in which case they can be included as negative costs in the numerator) or in the metric of the denominator. This is not an easy condition, as many projects have benefits in different fields. Free school meals, for instance, are supposed to have an effect both on learning and on health outcomes.

52. Under some conditions, then, the use of CERs is a full substitute for CBA. For this reason, evaluation based on CERs is considered a valid but nonetheless lower form of efficiency analysis in the scientific literature. This tool also satisfies the Drummond conditions, as costs and values constitute respectively the numerator and the denominator of the CER, and as CERs are designed for comparisons between at least two alternatives.⁶

53. Depending on which nonmonetary metric is selected, a CER can express any of the different levels of efficiency identified in Figure 2. To give an example from health: the effectiveness indicator can be patients enrolled in a health centre, patients treated at the health centre, patients cured, or life years thus saved. There is a natural priority ordering to

⁶ For an analysis of how rigorous effectiveness analysis can be expanded into CEA, with examples from the field of education, see Dhaliwal et al 2012.

these different calculations. The higher the indicator is situated in the logical framework, the better.⁷ If there is a contradiction between two rankings, the CER that is higher in the logical framework trumps the lower one. For example, health centre A, which has a lower cost per patient seen, is nevertheless less efficient than health centre B if the latter has a lower cost per life year saved. To give another example, in choosing among two road designs, the project design that achieves the best efficiency at outcome level will be preferred, even if it has a higher cost of supervision (lower efficiency at activity level) than a competing design.

54. It does not follow, however, that the use of lower-level CERs should be shunned. There is a trade-off, in that higher-level CERs tend to be more difficult to compute. The difficulty does not reside in the efficiency analysis per se, but in the underlying effectiveness analysis. In fact, the numerator expressing the costs will often be the same in both cases. It is the denominator, expressing the benefits in nonmonetary terms, that usually causes the trouble. Establishing the counterfactual becomes more complex as we proceed to the higher levels in the logical framework of a project—so much so that efficiency analysis may no longer be possible or worth the cost. However, any assumption that the transformation of outputs into intermediate and final outcome is so comparable that no ranking reversal can take place at higher CERs should not be made lightly, and should be argued.

Cost-utility analysis

55. Arguably the most sophisticated nonmonetary indicator used is *quality-adjusted life years* (QALY) saved. It was developed in the health sector as a nonmonetary metric at final outcome level. It is sometimes called cost-utility analysis (CUA). The indicator takes into account both the length of life saved and its quality. If two three-year-old children are saved through a medical treatment provided in a health clinic, but only the first has been reached in time to prevent lasting harm to its long term health, then the benefits are different. The first child will have a longer life expectancy than the second (quantity), but it will also have a healthier and more productive life (quality). The QALY captures both dimensions. A powerful advantage of the QALY indicator is that it allows ranking of health projects whose output and intermediate outcome indicators are different but whose final outcome indicator is the same. An example is a project that vaccinates children against measles and another project that offers treatment for HIV/AIDS. In both cases the final benefit is length and quality of life years saved, even if the indicators at lower levels are different. QALYs can be compared not only across interventions and sectors, but also across countries. Although QALYs are not easy to estimate, an impressive literature from the health sector provides many examples of how it can be done. The results of hundreds of painstaking empirical studies have led to a fairly good capacity to predict QALYs for a broad range of interventions. Once QALYs have been estimated, it is relatively straightforward to calculate CERs.

⁷ In the economics literature it is taken for granted that CEA aims for the highest level in the logical framework—intermediate outcome at a minimum, and final outcome where possible. CEA that is restricted to the output level is frowned upon (see, e.g., Drummond et al 1997: 102-103; Levin & McEwan 2001: 111-112). From this perspective it is of course odd that donors would be so focused on that lower level as to ban the use of the term *efficiency* at higher levels altogether!

Could we compare MDBs’ ability to lift people out of poverty?

56. What does our analysis indicate about the feasibility and relevance of trying to compare MDBs in terms of their record in lifting people out of poverty? In CEA and CUA the selection of the results metric determines how broad a range of projects can be ranked and compared. As indicated, cost per QALY is a metric that allows a comparison of almost all interventions in the health sector.⁸ It has been suggested (DFID 2014) that an even better CEA metric for the MDBs might be the cost per person lifted out of poverty (PLP). On the face of it, this is true: it would allow the use of CEA to compare almost all projects and programmes financed by the MDBs (except the EBRD). And it would avoid the controversy, and some of the thorny methodological challenges, of doing full CBA. It would in fact constitute a major innovation. But this measure has important conceptual as well as practical drawbacks, and for this reason we recommend not investing in it.

57. First, PLP is based on a binary choice. This means that the degree of improvement for those people who effectively cross some defined poverty threshold is not considered, only the fact that they have crossed it. Similarly, the improvements (or deteriorations) both for people who remain below and for people already above the poverty threshold are not captured at all. This means that a lot of the relevant consequences of a project are not considered.

58. Second, it would require a huge amount of rather precise information, and it is just not realistic to hope to estimate it in a reliable and truly comparable fashion across interventions, let alone across MDBs. Below (paragraph 68 onwards) we discuss an earlier failed effort to apply CBA at scale. PLP would be as cumbersome to calculate with any level of precision as full CBA, there would be huge difficulties of attribution, and we would be stuck with a rather flawed metric to boot. So the sobering answer to our question is negative: we cannot realistically compare MDBs portfolios on PLP..

Cost minimisation and yield maximisation

59. Two special cases of efficiency analysis are *cost minimisation* and *yield maximisation*. They are simpler to apply than CERs, for which both costs (numerator) and benefits (denominator) have to be estimated. Cost minimisation is an appropriate decision rule when comparing alternatives that have identical results, and yield (benefit) maximisation is an appropriate decision rule when comparing alternatives that have identical costs. Obviously this is not often the case, but if it happens, efficiency analysis is much simplified. Note that in all other cases cost minimisation amounts to mere cost cutting, which does not qualify as a VFM measure in the Drummond sense (see Box 3 above).

Other tools

60. An OECD discussion paper argues that VFM “does not need to be about monetising everything and applying cost-benefit or cost-effectiveness analysis”. And in another passage, CBA and CEA are described as tools “that may be relevant” at the level of individual projects and programmes, but “in the majority of cases a less ambitious assessment of value for

⁸ Even so, this does not mean that this would allow to rank MDB performance in the health sector just on this basis, for reasons presented in paragraph 81.

money will be more feasible” (Jackson 2012). This hints at the existence of other VFM tools outside those discussed in the previous paragraphs. But is this the case?

61. In a study for the German Ministry of Development Cooperation, BMZ, Palenberg 2011 identifies an extensive list of 15 different tools of efficiency analysis (see Box 4 below).⁹ Of these, he considers six to have high analytical power (level 2), in that they provide a robust overall assessment of the efficiency of an intervention, compared to alternatives. Seven are classified as having intermediate analytical power (level 1). They focus on some partial aspect of efficiency only, typically within a particular intervention, and this makes them incomplete and methodologically somewhat suspect. The remaining two tools receive a low score on analytical power (level 0), because they provide a qualitative but inevitably subjective expert assessment on certain dimensions pertaining to efficiency.

Box 4 A classification of efficiency tools

Level 2	Level 1	Level 0
CBA	Benchmarking of unit costs	Expert judgement
CEA	Follow the money	Specific efficiency questions
	Financial analysis	
	Stakeholder-driven approaches	
	Benchmarking of other indicators than unit costs	
Intuitive multi-criteria analysis	Stakeholder effectiveness and cost rankings	Specific efficiency questions
Scientific multi-criteria analysis		
Effects method	Stakeholder efficiency ratings	
CUA		

Source: Palenberg 2011: 92.

62. Of the 15 efficiency tools identified by Palenberg, 7 come straight from economics. Two of them are not of interest for our purposes. The Effects Method is a French variant of CBA (“méthode des effets”) that is as difficult to apply as CBA but may not always give the right answers, and has been argued to be methodologically flawed (Balassa 1976). And what Palenberg labels *financial analysis* is really CBA from the perspective of a particular actor or restricted group of actors (e.g., the government, a donor lending agency, a private implementing agency), whereas we require a societal perspective. Three others (CBA, CEA, and CUA) have already been discussed. The remaining two (benchmarking of unit costs, benchmarking of other than unit costs) will be briefly discussed further.

63. Among the remaining eight tools, two are versions—one simple, one complex—of multi-criteria analysis. It is difficult to imagine a convincing multi-criteria analysis of

⁹ As in this paper, Palenberg uses the broader DAC definition of efficiency, as also used in economics, rather than the narrow one that limits efficiency to the output level in the logical framework (Palenberg 2011, p8).

development projects that does not contain a CBA, CEA or CUA metric as one of the criteria, alongside, say, equity, environmental sustainability, and the like. Multi-criteria analysis is thus a natural complement to the economic efficiency tools already discussed, rather than an alternative. It is mainly used as a decision-making aid (ex ante analysis), but could in principle also be used as an instrument for accountability purposes (ex post analysis) if the intended audience can agree with the weights that are used to integrate the different dimensions into a single overall assessment.

64. The six that remain that are probably instances of the “less ambitious” but “more feasible” VFM assessment tools the OECD note is referring to (Jackson 2012). Half of them – stakeholder effectiveness and cost rankings, stakeholder efficiency rankings, expert judgement – may satisfy the Drummond conditions, but they are often subjective and difficult to compare across evaluations, and they may not be convincing VFM metrics, especially for accountability purposes. The other half – follow the money, stakeholder-driven approaches, and specific efficiency questions – may be useful for improving the management of a project being evaluated, but less for comparative learning across interventions, and they are even less likely to satisfy the Drummond conditions. In conclusion, although the Palenberg analysis starts with an impressive 15 efficiency tools, closer inspection reveals that it is difficult to identify rigorous VFM tools outside those developed in the economic efficiency literature.

65. This does not mean that such other tools, from level 1 and level 0 in the Palenberg classification (Box 4 above), are useless. Given the many theoretical and pragmatic difficulties involved in doing VFM at an aggregate portfolio level, to which the paper returns, some of those less robust indicators are valid additions. The argument in this paper is therefore not that such less-robust indicators do not have a place in VFM analysis of MDBs. It is rather that they have a number of weaknesses, and that an overall VFM assessment requires sufficient efforts to also include the more robust tools from welfare economics.

Unit costs benchmarking

66. Unit costs are calculated as a ratio of costs to some relevant unit (e.g., food, staff, beneficiary), and algebraically they resemble CERs. But the emphasis is not on comparing alternative strategies but on keeping down the costs of the alternative that has been chosen. Such a cost analysis does not inform us about the efficiency of development interventions.¹⁰ In fact, unit cost benchmarking does not pass the Drummond test that a good efficiency metric must consider both sides of the value and money equation and must at least compare the intervention with the counterfactual. Calculating unit costs and comparing them with some external benchmark can be a useful tool of cost containment, but it does not fully inform us about VFM.

67. Benchmarking can also be applied to other indicators than unit costs, such as the time for a loan to become effective, or the percentage of nonperforming loans in a portfolio. This tool is located in the lower levels of the logical framework (inputs and activities). Again

¹⁰ The author of the BMZ study interviewed a number of evaluation experts from the DAC network. This is how he summarised their views and his own position on unit cost benchmarking: “These experts saw too great a risk of explicit or implicit conclusions based on insufficient facts. Indeed, in our own review of evaluation reports that contained unit costs and other partial efficiency benchmarks, observations regarding overall efficiency of interventions were sometimes based on a fragment of the necessary information only” (Palenberg 2011: 44).

it is not a measure of VFM that satisfies the Drummond conditions, but it is nevertheless useful as an indicator of managerial efficiency.

Mainstreaming efficiency analysis: Lessons from the 1970s

68. The 1960s and early 1970s saw the publication of a number of path-breaking theoretical works on CBA applied to developing countries. These studies went beyond the traditional distinction between efficiency and equity that underpinned the famous Pareto and Kaldor-Hicks principles in economics.¹¹ Rather than ignoring equity, they instead postulated that a monetary gain to poor people should get a higher weight than the same gain to rich people. This principle was operationalised in CBA by a system of distributional weights. The same principle was also applied to intertemporal equity by selecting a social-time-preference discount rate as a function of economic growth, distributional weighting, and so-called pure time-preference. The opportunity cost of capital, traditionally expressed in the discount rate, was now captured in a new concept, the shadow price of investment, applied to all costs (and benefits) that displaced (or generated) investment.¹² Because in developing countries markets were also very much distorted, either because of underdevelopment itself (poorly integrated or non-existing markets) or because of government interventions, shadow prices were introduced for wages, the exchange rate, and many goods and services.

69. It took a while for all these considerations to be integrated in a consistent and convincing theory. The three leading publications (UNIDO 1972, Little & Mirrlees 1974, Squire & van der Tak 1975) set a new standard for efficiency analysis in poorer countries, and the underlying principles have withstood the test of time, even if the value judgements that they require are often contested. In fact, some of their novel features have found their way into the mainstream literature on CBA (e.g., Boardman et al 2011). These three publications are very similar in their treatment. The fact that the three books were published in such a short period of time and advertised their differences mainly reflected the intense intellectual competition that was going on. Interestingly, all three publications were sponsored by international agencies, one of them by the World Bank. Initially, the MDBs embraced this new approach, and several produced their own books and guidelines.

70. Of considerable importance to the present debate about VFM is what happened next. The introduction of CBA within the MDBs and the wider development community proved an almost perfect failure. The same organisations that had welcomed the new theory and had eagerly set out to apply it never succeeded in making it work in practice. The rapid rise and equally rapid fall of CBA was well documented by, among others, some of the original authors 20 years later (Little & Mirrlees 1991). Another 20 years later, the Independent Evaluation Group conceded the further decline both in quantity and in quality of CBA at the World Bank (IEG 2010). This does not mean that CBA and related forms of economic analysis are not applied at all. It arguably remains an important tool to inform decision

¹¹ The Pareto principle states that societal welfare improves when after an intervention some people are better off without anyone being worse off. The Kaldor-Hicks principles state that even if some people are worse off after an intervention, there is a welfare improvement if the winners could compensate the losers, even if they do not actually do so.

¹² In some of the publications, different concepts and parameters are used for discounting and the value of investment because of a change in the “numéraire”, but the final result is similar.

making in productive sectors, including at the MDBs. But it is seldom attempted in the full sense as proposed in the 1970s literature. Most applications are restricted to an ex ante financial CBA of individual projects, and there is little evidence of these tools being used to compare within or across sectors, let alone across countries. Outside the productive sectors, the situation is worse. Until recently it was used only to a limited extent in the education sector (Dhaliwal 2011). Discussing both the health and the education sectors, McEwan states that there is evidence that these tools are “under-utilised or imperfectly applied” to inform decision making (McEwan 2012: 190). On the more positive side, he notes a growing academic literature with applications and summative studies, a point we agree on and draw on in the recommendations.

71. What are the reasons for this decline? Here are a few that are relevant to the present discussion about VFM. For a start, the method proposed in the literature of the 1970s is very time-consuming and highly demanding in data. At the beginning of the project cycle there was often not enough time; in addition, the most important decisions—whether to go ahead with a project, or what its major design features would be—were typically decided before CBA was undertaken. At the end of the project cycle, not enough data were available to do a thorough post mortem. In addition, sophisticated efficiency-cum-equity CBA required so many assumptions regarding not only project parameters, but also “national” parameters such as shadow prices and distribution weights, the latter partly based on value judgements, that results could be easily manipulated. As CBA was often undertaken by staff who were closely involved in the project, they could make sure that the desired outcome emerged from their spreadsheet calculations. This, critics argued, imparted an air of futility to this use of CBA: in the best of cases it was irrelevant, justifying reasonable decisions already taken; in the worst cases, it served as a cover-up for erroneous decisions.

72. Furthermore, CBA operates under the assumption that resources are scarce and needs are almost unlimited. This is of course true, but that is not necessarily how it feels to MDB staff.

73. Another important factor hastened the decline of CBA: in the 1980s and 1990s practitioners increasingly realised that the counterfactual is more difficult to establish than the 1970s textbooks suggested. This problem arose in part from the fact that the textbooks were conceived for projects in the productive sector of the economy, such as industry, energy, mining, or transport. The counterfactual to the project was simply that no project would be undertaken, that resources would remain in their present use, and that, in the case of tradables, the output of the envisaged project would just be imported, or not be available at all. The counterfactual, in other words, could be extracted from the project feasibility study. It was almost taken for granted that the project under investigation would be undertaken by the public sector, so that private sector initiatives as counterfactual were largely ignored. But the perspective changed as a more market-oriented approach took hold of project decision-making at the MDBs in the 1980s and the scope of CBA was extended to social sectors such as health and education. It became increasingly clear that the identification of the counterfactual was much more complicated. The failure to apply rigorous tools of effectiveness analysis to measure the real counterfactual was seen as undermining the scientific value of any CBA built on it, however carefully it was executed in other respects (for an exposition of this argument in the area of health, see Hammer 1997, and in the area of education, see Jimenez & Patrinos 2008).

74. Against this background of both theoretical and pragmatic obstacles to the use of CBA, is it wise to advocate a more intensive use of this and related efficiency tools at the MDBs at this time? We think that both the theoretical and pragmatic arguments against its use must be addressed. MDBs should not apply the complicated form of CBA proposed in the 1970s. No MDB or other aid agency has ever worked with interpersonal distribution weights for policy decisions or for analysis. Our suggestion is to drop this dimension. As equity remains an essential part of the objective function pursued by the MDBs (with the possible exception of EBRD), they will have to address it in some other fashion. This could for instance be done by supplementing the efficiency metrics of CBA with information about the part of net benefits that accrue to the poor, or by choosing metrics in CEA that are sensitive to distribution.¹³ It is also possible to integrate equity in multicriteria analysis.

75. Intertemporal equity can be handled more easily through the discount rate, but it is controversial. And it may be argued that if interpersonal equity is left out, it is logical to do the same with intertemporal equity. The upshot of banning all equity considerations from CBA is that the discount rate will express the opportunity cost of capital, leading to much higher discounting of the future. This renders efficiency-only CBA particularly unsuited to address issues of ecological sustainability, such as investment to fight global warming. For such cases, a lower “social time preference” discount rate may be used, as for instance in Stern 2007.

76. Developing country economies are also much less distorted than they were in the 1970s, and the need for shadow prices has been correspondingly reduced (if not completely eliminated). This should make CBA simpler to implement. Alongside CBA, CEA is now being more used. CEA was not really considered important in the 1970s literature, but the subsequent expansion of efficiency analysis to the social sectors has made it an interesting tool. Finally, CBA and CEA need not be calculated for all projects or at every stage in their life cycle. They should be applied selectively.

77. The shift in emphasis between the 1970s and the present advocacy of efficiency tools is shown in Box 5 below.

¹³ “Food distributed,” for instance, is not sensitive to poverty, as people who do not really need food may nevertheless accept it and then sell it on the market. “Reduced undernutrition,” by contrast, is sensitive to poverty, as undernutrition is found mostly among the poor. Even if the better off who got hold of a free food ration were to consume it themselves, this would not lead to reduced undernutrition, as they were not suffering from any in the first place.

Box 5 Changing perspectives on efficiency tools

	1970s	2010s (VFM debate)
Stage in project cycle	Appraisal (ex ante)	Evaluation (ex post)
Equity dimension	Integrated	Kept separate
Market distortions	Major issue	Still key issue but somewhat less intractable
Application	Mainly productive sector	Sectors subject to market distortions and with public goods characteristics
Efficiency tool	Mainly CBA	CBA, CEA, CUA
Counterfactual	Not considered problematic	Recognised as the major challenge of rigorous evaluation

Aggregation works – up to a point

78. Many interesting syntheses and meta-evaluations are available from transport, irrigation, energy, health, nutrition, education, disaster prevention, and the like.¹⁴ They testify to the potential of aggregating rigorous efficiency analysis at the subsector level. The evaluation departments of some of the MDBs play an important role in performing such studies.

79. The strength of such aggregation depends crucially on the compatibility of the methodologies used in the underlying individual case studies. Even in sectors in which there has been much effort in this direction, much remains to be done. The best example is probably the health sector: a large number of high-quality studies using CBA and CUA are available for many countries and for a wide range of preventive and curative medical interventions. But whereas there is wide agreement among health economists that QALYs are an appropriate means to express health benefits across such applications, there are many ways in which this can be done in practice. This is because there are complex questions to be addressed—for instance, what discount rate to use, or how to weight the loss of life quality in case of physical or psychological impairment. The more unified the underlying methodologies are, the more forceful the aggregation results will be.

¹⁴ For instance, see Wei et al 2013, *Maternal vitamin D status and adverse pregnancy outcomes: A systematic review and meta-analysis*; UK Department for Transport 2014, *Meta Evaluation of Local Major Schemes: Transport-related Technical & Engineering Advice and Research – Lot 2 Roads*; or Saavedra and Garcia 2012, *Impacts of Conditional Cash Transfer Programmes on Educational Outcomes in Developing Countries. A Meta-analysis*.

80. MDBs could better harmonise their methodologies for CBA/CEA/CUA with a view to enhancing the power of meta-evaluations and syntheses. There are two dimensions to this. First, calculating the efficiency of interventions requires the application of concepts that are not project specific. Examples are discount rates to express the opportunity cost of investment, shadow exchange rates, shadow wages rates, corrections for artificial prices set by the public sector for electricity, and the like. In the literature of the 1970s, such 'national parameters' were supposedly set by a central planning authority and would be used for allocating resources in the public sector, whether by the government or international agencies providing loans or grants. But such national parameters are not usually available or reliable, and MDBs therefore tend to choose their own. Harmonisation here would mean that MDBs make an effort to streamline these parameters among themselves, so that their calculations become more readily comparable and meta-evaluations and syntheses become meaningful. The second dimension is that MDBs agree on some harmonisation of the methodologies used. If one MDB tries to include the environmental costs of road projects, and another does not, then obviously the results are not comparable. Harmonisation in this double sense would allow comparison within and across MDBs, at least in some sub-sectors, and for this reason be useful for learning, but also for accountability purposes. It would not be possible, on the basis of such aggregation, to rank MDBs in the way that some shareholders might like to see. But it would provide convincing evidence that, in the subsectors studied, MDBs provide VFM in the most complete sense of the term—that is, genuine improvements in human development at reasonable costs.

81. Unfortunately it is unlikely that such aggregation would be convincing at levels higher than subsectors. To come back to the health sector, an investment portfolio that is composed more of curative than of preventive health may show overall lower VFM, but it might make sense for a particular MDB to specialise more in curative health, given that it is only one of many actors in the health sector. Each country must find the right balance between preventive and curative health approaches, just as it must find the right balance between primary health care and the higher echelons in the pyramid, such as referral hospitals and university medical research facilities. But it does not follow that each MDB must of necessity invest in all of these in the same proportions; specialisation often makes perfect sense.

82. The same argument applies with greater force when we try to compare portfolios across different sectors and countries, as we would, in effect, have to do if we wish to rank MDBs for their overall VFM performance. MDBs work in different sectors and countries, and this fact seriously limits the meaningful aggregation from project to portfolio level. This paper argues that aggregation is possible and highly meaningful to the subsector level, but that beyond that level, we need additional tools to compare MDBs.

Accountability and learning

83. The meta-evaluations and syntheses discussed in the previous paragraphs are mainly performed for learning purposes. However, accountability and learning require different approaches and procedures. For instance, when learning is the purpose, so as to improve the functioning of an ongoing project, it is important to establish a relationship of trust between the evaluators and those being evaluated. But when stakeholders in a project believe that the evaluation is also performed for accountability reasons, they may not fully collaborate with an external team of experts.

84. Sampling of projects to be included in an evaluation is a standard procedure, whether the primary purpose is accountability or drawing lessons for future interventions. However, the principles behind the sampling are different in the two cases. For accountability, an independent audit office or a shareholder will be happy with evaluating only a sample of MDB projects, but will for obvious reasons insist that it, and not the MDB, determines the sample. As all projects could, at one time or another, be included in such a sample, their readiness for evaluation should be ensured at all times. Thus evaluation budgets may be spread too thinly, so that some superficial evaluation is performed for almost every intervention, but too few interventions are subject to rigorous evaluation. If, however, the evaluation is for learning purposes, in the sense of drawing lessons for future projects, the sampling may be done by the MDB itself, and on the basis not of some random (stratified) sample, but of a judgement as to where learning is most useful.

85. It is therefore important to strike a balance between these separate functions of evaluation. It is not easy to get it right. In a “client survey” organised by the World Bank’s Independent Evaluation Group (IEG) in 2014,¹⁵ respondents said that there was too much attention to accountability, to the detriment of learning. The more senior the World Bank staff, and the more familiar they were with IEG’s work, the larger they felt that the imbalance was. If evaluation is requested by a funding agency, rather than by an internal (even if autonomous) department such as IEG, the balance risks shifting even further towards accountability.

86. VFM has strong accountability connotations (e.g., it is associated with the activities of the UK national auditing office and the DFID pressure for multilateral VFM metrics). Most actors in the development community identify VFM with external accountability,¹⁶ and this is also how we interpret it here. Accountability is certainly important, and all the MDBs are concerned about accountability to their boards, to their clients from both the public and private sectors, and ultimately to the targeted beneficiaries. The MDBs’ managements have developed elaborate accountability and auditing procedures, as well as incentives to ensure that their corporate strategies are being implemented within the organisation and that personnel contribute to VFM.

87. But management should be equally concerned about learning, and the balance between accountability and learning should be respected. The tools described here for VFM, such as CBA, CEA and CUA, can serve both purposes. As learning has characteristics of a public good, it can be expected that learning evaluation will be underproduced, like all public goods. For effectiveness evaluation at the intermediate and final outcome levels, it has been argued that an “evaluation gap” has emerged “because governments, donors, official donors, and other funders do not demand or produce enough impact evaluations and because those that are conducted are often methodologically flawed” (Savedoff et al 2006). If this is true for effectiveness evaluation, there is no reason why the same political economy reasons would not also apply to efficiency evaluations. It is therefore reasonable to hypothesise that MDBs, like other donors, underproduce rigorous efficiency studies.

¹⁵ It can be accessed at <http://ieg.worldbank.org/Data/reports/ieg2014-client-survey-report.pdf>

¹⁶ This for example explains the tone in a recent OECD discussion paper (Jackson 2012).

The costs and benefits of doing project-level VFM analysis

88. Evaluations are expensive. They add to the overhead costs “that do not reach beneficiaries”. They have indirect costs as well as financial costs: for example, they require inputs from project staff and from beneficiaries, and sometimes this may affect the functioning of projects and programmes.

89. For smaller projects with benefits of limited size, the fixed cost of a rigorous evaluation may be many times the benefits the project conveys to beneficiaries. But even for large projects, a cost-benefit analysis of doing cost-benefit analyses (or CEA or CUA) may show that they are not worth doing. Projects that are fairly standard interventions about whose development effectiveness solid information has been gathered from similar projects elsewhere, such as scaling up of successful child vaccination campaigns, or primary education for girls, should not be subjected to the same deep evaluations every time, since we have reasons to be confident that outputs will translate into the expected intermediate and final outcomes. In such cases, evaluation may be restricted to establishing that timely outputs of sufficient quality have indeed been achieved at reasonable cost.

90. Critics have argued that a strong emphasis on VFM, with its accountability connotation, can lead to a bias in favour of projects with simple intervention logics and simple results indicators (e.g., using bed-nets against malaria) at the expense of more complex interventions (e.g., those emphasising local empowerment or institutional change in government) (Eyben 2013). Along the same lines, too much attention to VFM may lead to risk avoidance and an undue attention to short-term visible gains (Jackson 2012). It is therefore important that any MDB VFM metrics should not be applied in such a way as to introduce an undesirable bias in decision-making.

4. Applying VFM at aggregate level within and across MDBs

Overcoming the limitations of micro-oriented VFM tools

91. CBA and related VFM tools discussed in the previous section zoom in on the direct and indirect effects of an individual intervention, using what economists call a partial-equilibrium approach. Although CBA aims at capturing externalities and other economy-wide effects—using, among other things, shadow prices—certain systemic effects are beyond its grasp. These effects become visible only if one takes a sector or macro perspective. In what follows we distinguish three such effects that must somehow be taken into account in comparing VFM across MDBs: fungibility, systemic effects on the recipient government, and country portfolio effects.

Fungibility

92. Compared to budget support, funding projects (or programmes) is a form of donor conditionality: the funding is provided on condition that it is used for a particular project. Governments, however, do not always share the priorities of financing institutions, and may offer them projects they would otherwise finance themselves. Such fungibility creates fiscal

space that the government can use to increase government or private consumption, or to fund another investment project that it likes more but that is unpopular with donors. In the 1990s this became a major argument for abandoning project funding in favour of sector and macroeconomic support (World Bank 1998).

93. At present we are interested in how fungibility undermines the validity of microeconomic VFM calculations. VFM metrics (or at least those that satisfy the Drummond criteria) calculate a difference or ratio between the “value” (benefits, expressed either in monetary terms or in some physical unit) and the “money” (costs) of an intervention, compared to a counterfactual. We have noted how in the 1970s the literature on CBA for developing countries did not fully come to grips with the challenge of establishing a convincing counterfactual, because it took for granted that investing in productive projects was a task of the public sector. Since the 1980s this has become one of the major preoccupations in the scientific evaluation literature, and has led to a new approach to studying development interventions and a strong plea for the use of rigorous effectiveness analysis, in particular by using randomised control trials (see, e.g., Banerjee & Duflo 2011). But the new approach is essentially microeconomic, and it does not address the subtle political manoeuvring whereby the counterfactual of, say, an education project may well be a pay rise for the presidential guard. If such fungibility is taking place, the apparent conditionality of project financing is illusory.

94. The literature has given a lot of attention to donor-government fungibility and its consequences, but fungibility is just as likely to occur between donors. Suppose that several donors are funding projects in the health sector, and that they align with the government's sector investment plan. The projects have all been subjected to *ex ante* VFM analysis using CUA. There is quite a bit of competition among the funding agencies, and more money available than there are projects of acceptable quality in the investment plan. An MDB is lead donor in the sector and cherry-picks the project with the lowest cost per QALY. Other donors, lower down the pecking order, agree to fund projects that are still worth doing, but with a higher cost per QALY. So how does our expert assess the project the MDB is financing in this case? If she is very honest, she will admit that if the MDB had not selected this project, one of the other donors would have been very happy to take its place as a funder. In this sense, the net contribution of the MDB is more likely the last and least attractive of the projects selected from the government's list rather than the one it is actually financing.

95. If there is fungibility, whether of the donor-recipient or donor-donor variety, “what you see is not what you get”, project financing as a form of conditionality is an illusion, and attendant efficiency measures are off-target. But this conclusion has to be qualified, for it depends on why we are doing efficiency analysis. If the motivation is learning, it does not matter so much who is financing what. What matters is that the analysis helps us understand that certain types of interventions are worth doing, others not so well worth doing, and still others a waste of money. That is very important information for policymakers and funders alike. They can use it to make sure that the most promising activities get financed and implemented (and in the most efficient ways), by providing finance, but also by appropriate policy dialogue and technical assistance. However, when the motivation is accountability, fungibility is a damning criticism of the microeconomic focus of the VFM tools discussed in the preceding section. It should lead auditors, MDB board members, national parliaments and media to reject project-based VFM analysis in favour of sector and macroeconomic analysis of aid effectiveness and efficiency. Strangely enough, this is not what seems to

happen. Project and programme funding may often be an illusion, but it is a powerful one that most experts, opinion makers, audit agencies, and the public in donor countries are happy to entertain.

Systemic effects on recipient government

96. Development finance, especially in aid-dependent countries with poor governance, can easily create perverse effects. Some of these effects have been widely discussed (they were a major source of inspiration for the 2005 Paris Declaration). Sometimes these effects can be captured in micro-oriented VFM analysis. Consider the pressure from donor spending on the government budget. Funding agencies do not fund all the investment costs, much less the recurrent costs, of the projects they are supporting. Up to the 1980s, MDBs tended to fund financially profitable investment projects in the public sector; today, however, the projects and programmes they fund, often in the social sector, are typically not able to fully recover their costs. But less than full cost recovery implies that governments have to raise additional revenue, which in the absence of “optimal taxes” inevitably creates distortions in the economy. This can be captured in principle by applying a shadow price to marginal government income (Boardman et al 2011). Although such shadow pricing is not much done in practice, it could at least be envisaged. But other systemic effects elude even the most sophisticated CBA, and they give rise to what is sometimes described as the micro-macro paradox. If thousands of project-level evaluations exist and the majority suggest that development assistance is effective, why are these successes not visible at macro level, for example, when we run panel-data regressions with some development indicator (most often growth in per capita income) as the dependent variable, and assistance as one of the independent variables?

97. There is some discussion about the existence of the micro-macro paradox. Some argue, for instance, that most project evaluations are not worth the paper they are written on – not only are they methodologically flawed, but they are mainly commissioned by funding agencies and tend to be self-congratulatory by nature. Others argue on the contrary that it is the evidence for the macro part of the paradox that is untrustworthy, given the many methodological problems besetting cross-country regressions. Still others argue, on the basis of their own econometric analysis, that the effect of development assistance is positive and significant at the macro level, and the size of the effect is consistent with what can be expected from a realistic modelling of development assistance in the production function, which is in turn consistent with the positive signals coming out of project-level evaluations. Again this suggests that no micro-macro paradox need exist.

98. Nevertheless many researchers and practitioners believe that the paradox is real. Three plausible explanations are offered. The political explanation is that development assistance comes as an almost free resource to recipient governments, weakening their downward accountability to their own citizens and undermining already weak governance structures. The economic explanation highlights that development assistance gives rise to an inflow of foreign exchange, which may have a downward effect on the real exchange rate, penalising the traded sector of exports and import substitution, the most dynamic sectors of the economy. The third explanation says that isolated, unharmonised and unaligned interventions undermine the public sector, especially in aid-dependent countries with already weak public administrations. The three explanations are not mutually exclusive, and may to some extent reinforce each other. From the present perspective, the crucial point is

that, whichever explanation carries more weight, there is no way in which micro-oriented VFM analysis can capture such effects.

99. MDBs can mitigate these effects in several ways. They can make better use of political and political economy analysis in the policy dialogue; they can allow the Minister of Finance to modulate the absorption and spending of development assistance to monitor the real exchange rate; and they can more fully apply the Paris Declaration principles of harmonisation and alignment. These measures should help ensure that project results translate more readily into macro successes. The key observation here is that if MDBs do this, and thus improve the development value of their interventions, this should be reflected in their VFM metrics, but this does not appear to be happening. In some respects, the opposite is happening. Consider the choice of assistance modalities: these arguments should stimulate the MDBs to provide more aligned aid through budget support, when circumstances permit. Yet if they do so, it becomes much more difficult for them to show that they provide VFM, as the theory of change of such interventions is much more complex than the theory of change of most projects, the results take much longer to materialise, and it is difficult in any case for an MDB to prove that some of the results can be unequivocally attributed to it. The appropriate response is to complement micro-level VFM measures with macro analysis of the MDBs’ assistance modalities and strategies.

Country portfolio considerations

100. If MDBs only care about getting the best possible scores on micro-oriented VFM metrics, they should invest in member countries that have large numbers of poor people and relatively good governance. Such countries are where the fastest progress in development outcomes can probably be achieved at the lowest cost. Along these lines it has been argued that MDBs should focus on poor people, wherever they live, rather than on poor countries, and on interventions that have the highest VFM. In its extreme form, this would mean that MDBs should avoid getting working in fragile member countries with poor governance. But it is doubtful that is the best strategy. Fast-growing countries will do well without development assistance, and the comparative advantage of development assistance may be exactly in those fragile, poorly governed countries that are not able to profit from the possibilities that the international economy is offering (as argued, for example, by Collier 2007). The argument is not settled, but there are good reasons why MDBs, like other aid agencies, should not only focus on where the poor live, but also on some country characteristics (Kanbur & Sumner 2012); and if country portfolio considerations have some role to play in allocating resources across countries, then microeconomic tools such as CBA or CEA cannot be the whole story. We can conclude, therefore, that micro-oriented VFM must be complemented by a number of more macro-oriented considerations.

Other tools for aggregate VFM analysis

101. There is a very different approach to VFM assessments of MDBs that we have not discussed yet: cross-country econometric analysis. There is a vast literature on development effectiveness using this very statistical approach (see, e.g., Burnside & Dollar 2000, Clemens et al 2012, Easterly et al 2004, Rajan & Subramanian 2008, Roodman 2015). Why not use this approach to compare MDBs? It would run as follows. First we would select a development outcome at final outcome level, for which we have more or less reliable and ample data over time for the countries where MDBs operate. Growth in per capita income

would be an obvious candidate. Variables that are more poverty-related—for instance, involving the proportion of people living in extreme poverty—would arguably be even more appropriate, but data may not be of good enough quality or their coverage complete. We would then try to explain statistically the variation in this dependent variable over time between countries, by invoking a suitable number of explanatory variables, such as initial income level, geography, colonial history, economic policy, ethnic heterogeneity, and the like. To this we would then add a variable for development assistance received. The significance and sign of this variable then tells us whether variation in development outcomes is positively associated with development assistance. If we have convincing arguments to interpret this correlation as a causality running from financing to development outcome, we would have a direct aggregate measure of the overall effectiveness of development assistance. With further analysis we can even hope to extend the interpretation from effectiveness to efficiency. The beauty of this approach is that you do not have to study VFM at the micro level and then painstakingly aggregate, only to discover that this leaves out many things that you subsequently have to bring into the analysis.

102. This is in fact what the literature tries to do for aid from all sources, but it is controversial for a number of methodological reasons, and no clear conclusions have emerged so far. If we would like to replicate this analysis, but this time focused on MDBs, we would have to break down the development assistance variable into financing from every individual MDB we are interested in, plus a category for aid from other sources. In practice this would result in seven variables for the group of MDBs that commissioned this report. In addition, individual MDB interventions are often negligible compared to the borrower's economy, especially in the larger countries. It is not difficult to predict that the resulting estimates would be even more unstable and even less significant than in the present debate on development assistance. Although additional econometric work may turn up interesting insights through better data and more sophisticated statistical techniques, for the moment this is not a very promising course for studying the VFM delivered by the MDBs.

103. The analysis up to this point suggests two different approaches to VFM analysis of MDBs.

- Rigorous project-level efficiency analysis suitably aggregated through meta-evaluation and synthesis evaluations. This is the most convincing type of VFM analysis in two regards: it studies VFM at intermediate and final outcome levels, and it is scientifically robust. But such exercises cannot be performed up to the level of aggregation that may be required, because MDBs have heterogeneous portfolios that span many sectors, intervention modes and countries. Not only would the cost of bottom-up aggregation up to that level be prohibitive, but such aggregate VFM would also be flawed because project-level efficiency analysis does not capture all the relevant considerations.
- A more institutional analysis of MDB interventions would look into the issues of fungibility, systemic effects on governments, and strategic portfolio considerations. Here there is no one toolbox that is available, and the analysis would be mixed, using a variety of quantitative and qualitative methods.

104. This paper argues in favour of adding a third VFM component that looks at MDBs' operational efficiency at aggregate level. One metric suggested by an MDB is “percentage of problematic projects”—a metric that is directly captured at aggregate level. Strictly speaking

it is not a VFM metric in the methodological sense that we proposed for project-level analysis: although it says something about problematic projects’ benefit-cost ratio (presumably unsatisfactory), it is not clear on what basis such labels are given or to what extent every project falls short. But it may nevertheless be considered relevant information. Even though some failures are inevitable in the risky environment where MDBs operate, an unusually high percentage of problematic projects suggests that the MDB in question, compared to its colleagues, either does not do enough to stop doubtful projects at the appraisal stage, or does not do enough to resolve implementation problems or phase projects out once they turn problematic.

105. Many similar indicators exist, and MDBs already collect and report on some of them; indeed, some are even brought together in tables that compare MDBs (see Box 6 below). Most of these are at input level (e.g., staffing levels, employee costs), activity level (e.g., consultancy costs) or output level (e.g., unit costs), and all the limitations discussed above apply. In principle one can imagine an MDB doing a wonderful job at managing projects and programmes that do not bring lasting development outcomes, or do so only at exorbitant costs. Therefore these are not adequate VFM indicators if taken in isolation. But in combination with the two other building blocks proposed—efficiency metrics at outcome level for suitable samples of projects, aggregated at sub-sector level; and an analysis of institutional effects of MDB interventions not captured by project level tools—these operational efficiency indicators bring an additional dimension.

Box 6 MDBs’ Common Performance Assessment System

Since 2005 the MDBs have had a framework for common reporting of results—the Common Performance Assessment System (COMPAS) (<http://www.mfdr.org/Compas/index.html>).

COMPAS focuses on measuring the MDBs’ capacities to apply and improve operational processes toward achieving results on the ground. As a self-reporting exercise, COMPAS aims to measure MfDR capacity and progress consistently through the analysis of key performance indicators consolidated into a concise and convenient format, and comparable within each MDB. COMPAS is not designed to make direct comparisons across institutions, although the matrix format of the report does provide opportunities for the MDBs to learn from one another.

COMPAS does not incorporate VFM indicators as such (in the terminology of this paper, its focus is on effectiveness, which is one of the twin pillars of VFM). However, MDBs considering common VFM reporting will want to draw on this experience of reporting performance, consider whether any of the existing results metrics are relevant also to VFM, and seek to avoid unnecessary duplication of reporting processes.

A VFM matrix

106. In Box 7 below we sketch an outline of a VFM reporting format that reflects the foregoing analysis. The matrix is composed of four horizontal layers. The second highest layer is about programme delivery, arguably the most important component and the key ingredient of VFM analysis, and is directly derived from the analysis presented in this paper. CBA, CEA and CUA take central stage here, but other tools can also be included, such as the scoring systems that many MDBs have developed to assess the overall success rate of their interventions. In addition to developing appropriate versions of the efficiency toolbox, the MDBs will have to promote its use throughout the organisation and set up a quality support

and quality control function. Finally it will be important that MDBs sit together to harmonise the tools and metrics they use so as to facilitate comparison between them, again for both accountability and learning purposes. It is one of the legitimate requests of DFID 2014 that some of the metrics be common across the MDBs. This goes beyond all MDBs agreeing, for example, to apply CEA in certain sectors. It also means that each of them uses this particular tool in the same way. For instance: are costs incurred by beneficiaries (e.g., financial contributions, time) taken into account in the same way; is discounting applied to the non-monetised results indicator, and is it the same discount rate across MDBs?

107. The top layer in Box 7 is about the systemic effects that we have discussed, whose existence in our view warrants a complement to the micro-perspective of the traditional VFM tools. It follows the distinction between fungibility, national effects on government, and country portfolio considerations and provides a tabular synthesis of the argument developed.

108. The two bottom layers in Box 7 concern issues of MDB management that go beyond individual programme delivery and provide additional insights by looking at management processes. MDBs together spend billions of dollars on administrative costs every year. DFID 2014 proposes that the different components of administrative costs, such as salaries, consultancy services, travel, and training, be made more comparable across the MDBs and that the drivers of these costs be better understood and justified. A few metrics are presented in Box 7 for illustrative purposes, but they merit more reflection and discussion among MDBs. Most of these metrics will probably relate to the cost side or another partial aspect, and thus they do not constitute full VFM tools. For instance, they will not satisfy the Drummond conditions. Nevertheless, it may make sense to compare some of these metrics across the MDBs and try to understand whether they can be interpreted as indirect indicators of efficiency at activity level. In DFID 2014 this analysis attracts much attention, somewhat to the detriment of programme delivery itself and of the macro effects, both of which the present paper emphasises.

109. In Box 7 we distinguish between two dimensions of MDB management. Under portfolio/use of capital we suggest retaining indicators that are relevant to accountability but may not be well captured in the detailed study of programme delivery in the next higher layer of the matrix. The percentage of problematic projects is an example. As we argued before, we suggest that the sample for assessing programme delivery be drawn so as to maximise learning. The consequence is that the share of problematic projects that derive from the next higher layer of the analysis may not be representative of the total population. As data on the percentage of problematic projects will be available from MDB management, and as it seems not unreasonable for the MDBs to agree amongst themselves to define what “problematic” means and how to assess it, such an indicator conveys useful additional information. Under administration, the lowest layer, the emphasis is on administrative costs and their tracking. These are obviously important, and they should certainly be included, provided the cost categories used by the MDBs are sufficiently comparable. Also, care should be taken that considerations of cost minimisation and other activity-level concerns do not come to dominate the VFM approach. It is for this reason that it is probably useful for MDBs to develop a matrix along the lines of Box 7, in which the different aspects of VFM are kept in balance.

Box 7 A comprehensive VFM assessment matrix

Level	Focus	Tools / metrics (not exhaustive)
Systemic effects	➤ portfolio strategy	➤ thematic and sectoral studies on effects of MDB programmes on strategic objectives such as state strengthening, global public goods
	➤ national effects	➤ Paris Declaration monitoring indicators
	➤ fungibility (not a major issue for learning, but crucial for MDB accountability)	➤ macroeconomic and econometric analysis ➤ political economy analysis
Programmes	<ul style="list-style-type: none"> ➤ focus is on accountability, but compatible with learning ➤ a core part of the indicators presented here satisfy the Drummond conditions 	<ul style="list-style-type: none"> ➤ appropriate versions of CBA, CEA and CUA developed and promoted for wider use within the MDBs ➤ other tools, such as internal scoring of interventions, if performed by a (relatively) independent department, and if scoring procedure minimises subjectivity ➤ quality support and control department ➤ harmonisation of tools and metrics across MDBs is necessary to facilitate comparison
Portfolio / use of capital	➤ portfolio management (tools and metrics at aggregate level that are complementary to those at programme level and are project-based but do not satisfy Drummond conditions)	<ul style="list-style-type: none"> ➤ % of problematic projects ➤ % of projects eligible for cancellation
Administration	➤ institutional management (tools and metrics at aggregate level that are mainly concerned with activity and output levels)	<ul style="list-style-type: none"> ➤ overheads as % of spending ➤ tracking of employee costs ➤ tracking of consultancy costs

5. Recommendations

Main recommendations

110. Here we gather the main policy recommendations that follow from our detailed analysis. These recommendations should not be interpreted as a plea for additional investment in VFM at any cost. The principle is rather that such analysis should be expanded as long as its benefits plausibly exceed its costs. These benefits are the increased effectiveness of the funds channeled through the MDBs in terms of the ultimate objectives of sustainable human development. MDBs should welcome debate about VFM, but should insist on a rigorous definition of VFM that emphasises the analysis of value, not just cost, and focuses as much as possible on impact-level benefits. VFM is a term that needs to be used carefully and with attention to the relevant underlying economic concepts, based on the relationship between efficiency and effectiveness. There is scope for MDBs to work together to standardise techniques that are rigorous without being overambitious.

111. At the same time, both pragmatic and theoretical arguments support the use of additional metrics. These may include efficiency metrics at lower levels than impact (in which case they must be used carefully so as not to result in suboptimal decisions that focus too much on cost and not enough on value). They may also include process indicators that provide useful management information at organisational and portfolio levels as well as at project level.

112. It is important not to expect efficiency metrics to resolve the whole VFM debate. They are most robust when considering similar operations in similar contexts. Other DAC criteria, particularly relevance, should be applied in tandem. (So, for example, rather than seek direct comparison between operations in fragile and non-fragile contexts, the relevance criterion may be used to establish whether MDBs are right to operate in fragile contexts, with efficiency (VFM) comparisons then being made between similar fragile-context operations.)

113. **Recommendation 1:** *The “value” in VFM should be understood as referring to improvement of development outcomes for the final beneficiaries, in particular the poor, measured as objectively and accurately as possible. There is an accountability connotation, but VFM should not be assessed only in terms of the way such outcomes are perceived or assessed by northern taxpayers or governments. To avoid any misunderstanding in this respect, the MDBs should use a broad definition of efficiency, as used in economics. Such a definition is not incompatible with the DAC terminology.*

114. It is now generally accepted that assessing effectiveness is one of the most challenging parts of any evaluation, because it requires a comparison of what is happening (which can be directly observed in the projects during implementation) and a counterfactual of what would have happened in the absence of a project (which by definition cannot be directly observed). Evaluation departments are increasingly relying on experimental and quasi-experimental designs to test whether interventions work. VFM analysis builds further on effectiveness analysis. If effectiveness has not been properly assessed, no efficiency analysis will be convincing. A rigorously performed effectiveness evaluation provides the basis for an equally rigorous efficiency analysis. For CEA, it is only the cost side that is left to estimate. For CBA, the outcome/impact indicator also has to be monetised. Compared with the effort that has gone into outcome/impact evaluation, the additional work required is

relatively minor. **Recommendation 2:** *In evaluation strategies, MDBs should therefore seek better complementarity between the rigorous impact evaluations they increasingly organise and VFM analysis.*

115. There are many possible tools for efficiency analysis, and thus for VFM. Most, but not all, are quantitative, and some are methodologically more robust than others. The more qualitative tools, such as expert judgements, are very important; indeed, they are the main evaluation instrument donors use for interim evaluations, in which lessons are drawn for the project in question. But when evaluations are performed for future learning, we need methodologically robust, quantitative tools to aggregate across projects and perform meta-evaluation. When performing efficiency analysis for accountability purposes, and the evaluation applies to a sample of projects, or to a sector or subsector, or to a whole country portfolio, we similarly need a methodologically robust approach. Hence **Recommendation 3:** *MDBs should make sure that the core VFM measures they use for individual or samples of projects satisfy what we have called the Drummond conditions, i.e. that both the money (or cost) side and the value (or benefit) side are considered, and that every intervention is at least compared with a counterfactual.* Measures that do not respect these minimum conditions are not satisfactory for accountability purposes when performed in isolation. Nor are they an adequate basis of learning for future interventions. They can nevertheless be used for additional information, when they come in addition to some robust VFM metric.

116. It is important that analyses undertaken by different MDBs are, to the extent feasible, mutually consistent and comparable. Hence **Recommendation 4:** *MDBs should establish appropriate quality controls to make sure that efficiency analysis is carried out according to minimum methodological standards and with sufficient care. To facilitate comparison across MDBs, for both accountability and learning purposes, MDBs should endeavour to harmonise their approaches, as in the selection of discount rates. For CEA and CUA, some collaboration among MDBs to harmonise the selection and precise calculation of results metrics would also be important and useful.* To give an example, for calculating quality-adjusted life years (QALYs), for which a substantial literature and practice already exist, there is quite some discussion on how to calculate them precisely: whatever the merits of fine distinctions, adopting a common approach would have benefits in supporting meta-analysis.

117. Donor agencies have a tendency to lean too much towards the accountability side, and the present emphasis on VFM, by audit agencies in donor countries and the EU, probably further increases this gap. Yet the good news is that methodologically sound efficiency analysis can serve both the learning¹⁷ and accountability functions of evaluation. **Recommendation 5:** *MDBs should carefully reflect on how the present emphasis on accountability could undermine the learning side of evaluation.* This perverse effect could occur because too many projects are subjected to superficial efficiency analysis that is itself often based on unreliable effectiveness analysis, all for the sake of satisfying the demands of shareholders, audit agencies and public opinion. If VFM can be done on the basis of sampling, and projects selected in such a way as to address learning issues, then this conflict need not arise. In a note on measuring VFM at the MDBs, DFID 2014 makes several thoughtful comments on this that suggest that at least this donor will be sensitive to such

¹⁷ In the sense of ex post learning for future projects.

arguments. For instance, rather than monitoring the *results* of CBA studies, MDBs might offer enough information to monitor the *use* of CBA, which would leave MDBs in the driver's seat in determining which projects to subject to such analysis.¹⁸

118. The project-level VFM tools we have reviewed (CBA, CEA and CUA in particular) have a solid theoretical justification. But their grounding in microeconomics also carries some important limitations. Because of fungibility between MDBs and government, and between MDBs and other donors, VFM assessments may focus on the wrong intervention. VFM assessments also fail to capture the unwanted negative effects donor projects may have on the government, the administration, or the exchange rate, especially in aid-dependent countries with poor governance. Finally, given country portfolio considerations, MDBs may be justified in not limiting themselves to maximising VFM as measured at the micro level, but including other objectives such as a focus on fragile states. MDBs should be concerned about these issues (even if their funders may not be) because they influence development results, whether this is captured in elaborate metrics or not. Hence **Recommendation 6:** *MDBs should study the sector and macro effects of their interventions alongside project-specific efficiency analysis, and they should report on all these effects. They can do so by making use of some of the DAC criteria. Relevance, for instance, can be invoked to capture donor portfolio considerations, and sustainability can be likewise invoked to capture fungibility and some of the unintended negative effects of isolated projects on government.*

119. While MDBs should insist on rigorous approaches to efficiency analysis (and should be correspondingly frugal in the use of the term *value for money*), they should also recognise that regular reporting for management purposes needs to draw on a wide set of metrics, including but not limited to rigorous VFM analysis. These metrics can draw on partial efficiency analysis (e.g., monitoring costs of inputs) as well as on process indicators (projects delivered on time, application of efficiency criteria in selection and monitoring of projects, etc.). To avoid the potential distorting effects of taking such metrics in isolation, it would be valuable to develop a coherent schema that balances different levels and types of analysis and metric. **Recommendation 7:** *MDBs should collaborate to develop a schema of VFM and related performance metrics that fit their respective institutions, so as to ensure that they maintain a proportionate view of VFM issues while seeking to maximise learning and the quality of management decisions.* This could be a way for MDBs to seize the initiative to improve their VFM-related reporting, and do so in a way that is methodologically sound, does not distract from the learning function of evaluation, and avoids the perverse incentives that externally imposed rules may have. Finally, **Recommendation 8:** *Such a schema should be designed to build as much as possible on existing MDB reporting frameworks, and to support internal decision-making, accountability and improvements over time as much as inter-MDB comparisons.*

¹⁸ A more general point is that all the proposed metrics in Box 7 need to be deployed with care. Strong incentives to report the “right” results can undermine the credibility of the data (a point that applies as much within as across MDBs).

References

short ref	full ref
Balassa 1976	Balassa, B. (1976), “The ‘Effects Method’ of Project Evaluation.” <i>Oxford Bulletin of Economics and Statistics</i> , 38(4): 219-31.
Banerjee & Duflo 2011	Banerjee, A.V., Duflo, E. (2011), <i>Poor Economics. A Radical Rethinking of the Way to Fight Global Poverty</i> , Public Affairs Books.
Boardman et al 2011	Boardman, A.E., Greenberg, D.H., Vining, A.R., Weimer, D.L. (2011), <i>Cost-Benefit Analysis – Concepts and Practice</i> , fourth revised edition, Pearson.
Burnside & Dollar 2000	Burnside, C., Dollar, D. (2000), “Aid, Policies, and Growth.” <i>The American Economic Review</i> 90:847–68.
Clemens et al 2012	Clemens, M. A., Radelet, S., Bhavnani, R. R., and Bazzi, S. (2012), “Counting Chickens when they Hatch: Timing and the Effects of Aid on Growth”, <i>The Economic Journal</i> , 122: 590–617.
Collier 2007	Collier, P. (2007), <i>The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About it</i> , Oxford, Oxford University Press.
DAC 2002	DAC (2002), <i>OECD Glossary of Key Terms in Evaluation and Results-Based Management</i> , Paris, OECD Development Assistance Committee (2002, reprinted 2010). http://www.oecd.org/development/peer-reviews/2754804.pdf
Devarajan et al 1997	Devarajan, S., Squire, L., Suthiwart-Narueput, S. (1997), <i>Reviving Project Appraisal at the World Bank</i> , Policy Research Working Paper 1496, Washington D.C., World Bank.
Dhaliwal 2011	Dhaliwal, I. (2011), <i>Comparative Cost-effectiveness Analysis to Inform Policy in Developing Countries</i> , Abdul Latif Jameel Poverty Action Lab (J-PAL), Massachusetts Institute of Technology (powerpoint)
Dhaliwal et al 2012	Dhaliwal, I., Duflo, E., Glennerster, R., Tulloch, C. (2012), <i>Comparative Cost-effectiveness Analysis to Inform Policy in Developing Countries</i> , Abdul Latif Jameel Poverty Action Lab (J-PAL), Massachusetts Institute of Technology.
DFID 2014	DFID (2014), <i>Measuring VFM at the MDBs</i> , unpublished document.
Drummond et al 1997	Drummond, M.F., O’Brien, B., Stoddart, G.L., Torrance, G.W. (1997), <i>Methods for the Economic Evaluation of Health Care Projects</i> , second revised edition, Oxford University Press.
Easterly et al 2004	Easterly, W., Levine, R., Roodman, D., 2004. “Aid, Policies, and Growth: Comment.” <i>American Economic Review</i> , 94(3): 774-780.

short ref	full ref
ECG n.d.	"A Note on 'Impact' and 'Impact Evaluation' in the Public Sector GPS", an explanatory annex to the "Big Book on Evaluation Good Practice Standards" published by the Evaluation Cooperation Group (a collaboration among the major development banks), and available at https://wpqr4.adb.org/LotusQuickr/ecg/PageLibrary48257B910010370B.nsf/h_9DF721C00F70E93948257B9500303684/DFC7E13EBD11975B48257B950034FE96/?OpenDocument
Eyben 2013	Eyben, R. (2013), <i>Uncovering the politics of "evidence" and "results". A framework paper for development practitioners</i> , available at: www.bigpushforward.net
Hammer 1997	Hammer, J. (1997), "Economic Analysis for Health Projects", <i>The World Bank Research Observer</i> , 12 (1): 47-71.
ICAI 2011	<i>ICAI's Approach to Effectiveness and Value for Money</i> , Independent Commission for Aid Impact Report 1, November 2011.
IEG 2010	IEG (2010), <i>Cost-Benefit Analysis in World Bank Projects</i> , Washington D.C., World Bank.
Jackson 2012	Jackson, P. (2012), <i>Value for Money and International Development: Deconstructing Myths to Promote a More Constructive Discussion</i> , OECD DAC, Paris.
Jimenez & Patrinos 2008	Jimenez, E., Patrinos, H.A. (2008), <i>Can Cost-Benefit Analysis Guide Education Policy in Developing Countries?</i> Policy Research Working Paper 4568, Washington D.C., World Bank.
Kanbur & Sumner 2012	Kanbur, R., Sumner, A. (2012), "Poor Countries or Poor People? Development Assistance and the New Geography of Global Poverty", <i>Journal of International Development</i> , 24: 686-695.
Levin & McEwan 2001	Levin, H.M., McEwan, P.J. (2001), <i>Cost-Effectiveness Analysis</i> , second edition, Sage Publications.
Little & Mirrlees 1974	Little, I. M. D., Mirrlees, J. A. (1974), <i>Project Appraisal and Planning for Developing Countries</i> , New York, Basic Books.
Little & Mirrlees 1991	Little, I. M. D., Mirrlees, J. A. (1991), "Project Appraisal and Planning Twenty Years on", in <i>Proceedings of the Annual Conference on Development Economics 1990</i> , p. 357-395, Washington D.C., World Bank.
McEwan 2012	McEwan, P.J. (2012), "Cost-effectiveness analysis of education and health interventions in developing countries", <i>Journal of Development Effectiveness</i> , 4(2): 189-213.

short ref	full ref
OECD 2015	<i>Building the Basics, Value for Money in Government</i> , Paris http://dx.doi.org/10.1787/9789264235052-en
Palenberg 2011	Palenberg, M. (2011), <i>Tools and Methods for Evaluating the Efficiency of Development Interventions</i> , Evaluation Working Papers, Bonn, Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung.
Rajan & Subramanian 2008	Raghuram G. Rajan. R.G., Subramanian, A. (2008), “Aid and Growth: What Does the Cross-Country Evidence Really Show?” <i>The Review of Economics and Statistics</i> , 90(4): 643–665.
Renard 2014	Renard, R., (2014), <i>Why Aid Agencies Neglect the Efficiency Dimension in Evaluation</i> , presented at European Evaluation Society Dublin Conference 1-3 October 2014.
Roodman 2015	Roodman, D. (2015), “A Replication of ‘Counting Chickens When They Hatch’ (<i>Economic Journal</i> 2012)”, <i>Public Finance Review</i> , 43(2): 256-281.
Savedoff et al 2006	Savedoff, W., Levine, R., Birdsall, N. (co-chairs) (2006), <i>When Will We Ever Learn? Improving Lives through Impact Evaluation - Report of the Evaluation Gap Working Group</i> , Washington D.C., Center for Global Development.
Squire & van der Tak 1975	Squire, L., van der Tak, H.G. (1975), <i>Economic Analysis of Projects</i> , Baltimore and London, Johns Hopkins Press for the World Bank
Stern 2007	Stern, N. (2007), <i>The Stern Review on the Economics of Climate Change</i> , Cambridge, Cambridge University Press.
UNIDO 1972	UNIDO (1972), <i>Guidelines for Project Evaluation</i> , New York, United Nations.
World Bank 1998	World Bank (1998), <i>Assessing Aid – What Works, What Doesn’t, and Why</i> , Washington D.C., Oxford University Press and World Bank.

Acronyms and abbreviations

BCR	benefit-cost ratio
BMZ	Federal Ministry for Economic Cooperation and Development (Germany)
CBA	cost-benefit analysis
CBR	cost-benefit ratio
CEA	cost-effectiveness analysis
CER	cost-effectiveness ratio
COMPAS	Common Performance Assessment System
CUA	cost-utility analysis
DAC	Development Assistance Committee (of the OECD)
DFID	Department for International Development (UK)
EBRD	European Bank for Reconstruction and Development
EC	European Commission
ECCG	Evaluation Cooperation Group
ICAI	Independent Commission for Aid Impact
MDB	multilateral development bank
NPV	net present value
OECD	Organisation for Economic Co-operation and Development
PLP	person lifted out of poverty
QALY	quality-adjusted life year (gained)
UN	United Nations
VFM	value for money