Image by LumenSoft Technologies

SECOND EDITION

Assessing Value for Money: the Oxford Policy Management Approach

Julian King, Daniel Wate, Esther Namukasa, Alex Hurrell, Frances Hansford, Patrick Ward and Shiva Faramarzifar

Oxford Policy Management

JK&A

Finally, an approach to Value for Money that breaks free of the 'here's the formula' approach and instead emphasises the importance of thoughtful and well-evidenced evaluative reasoning. Combining an equity lens with insights and perspectives from diverse stakeholders helps us understand the value of different constellations of outcomes relative to the efforts and investments required to achieve them. This step-by-step guide helps decision makers figure out how to answer the VfM question in an intelligent way when some of the most valuable outcomes may be the hardest to measure—as they so often are.

Dr E. Jane Davidson, author of Evaluation Methodology Basics (Sage, 2005); Director of Real Evaluation LLC, Seattle; 2019 winner of the AEA Lazarsfeld Award for Evaluation Theory.

Oxford Policy Management Limited
Registered in England: 3122495

Level 3, Clarendon House
52 Cornmarket Street
Oxford, OX1 3HJ
United Kingdom

Tel: +44 (0) 1865 207 300
Fax: +44 (0) 1865 207 301
Email: admin@opml.co.uk
Website: www.opml.co.uk
Twitter: @OPMglobal
Facebook: @OPMglobal
YouTube: @OPMglobal
LinkedIn: @OPMglobal

Foreword

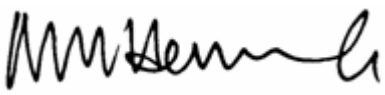
Aid budgets are under intense scrutiny; accordingly, aid agencies and development practitioners face a growing need to demonstrate Value for Money (VfM). This is appropriate and to be welcomed. However, until recently, this need had not been matched with appropriate methods to support meaningful VfM evaluation – with the prevailing tendency being to rely on readily quantifiable measures, even when those do not capture the most important aspects of the change being pursued.

In 2016, Oxford Policy Management (OPM) teamed up with Julian King, an evaluation specialist, who worked with staff from across the company to develop a robust and distinct OPM approach to assessing VfM. Combining contemporary theory and practice from evaluation and economics, our approach provides a system for aligning multiple methods and tools to support transparent, evidence-based judgements about how well resources are being used and whether enough value is derived to justify the investment. The approach was successfully piloted in two UK Aid programmes: MUVA, a women's economic empowerment programme in Mozambique, and the Sub-National Governance (SNG) Programme in Pakistan.

We have now applied this approach to numerous development projects and programmes spanning a range of clients, countries, sectors, and budgets. It has been well received by our clients (both funding agencies and partner governments) and project teams alike, who particularly appreciate the use of explicit evaluative reasoning. This involves developing definitions of what acceptable, good, and excellent VfM looks like in the context of each project. Critically, these definitions are co-developed and endorsed upfront, in advance of implementation and before evidence is gathered, which provides an agreed, explicit, and transparent basis for making judgements.

Responding to high demand, in 2018 we published a short, practical guide on our approach to VfM, which detailed the underlying conceptual framework and set out the practical steps required to apply it, including how these can be integrated with project evaluations and routine management activities. The guide has lived up to our hope of making a valuable contribution to strengthening the accountability, learning, adaptation, and impact of development interventions. Our approach has not only been used by our clients, but also by other consulting firms, government agencies, bilateral and multilateral organisations, and non-governmental organisations worldwide.

It now gives us great pleasure to publish the Second Edition of the guide, building on our more recent experience in using the approach in an increasingly diverse range of settings.



Mark Henstridge

Chief Executive Officer
OPM

Executive summary

This document offers practical guidance for assessing the Value for Money (VfM) of government- and donor-financed programmes and policy interventions. In line with OPM's focus and mission, it has been predominantly applied in the international development sector, but the approach upon which it is based is also used in the context of domestic public policy and programmes.[1]

There is increasing scrutiny on VfM in international development, but a lack of appropriate methods to support its assessment. There is a risk of reaching invalid conclusions if VfM evaluation is tied to a narrow set of indicators devoid of any evaluative judgement—for example, by emphasising the most readily quantifiable measures rather than the most important (but harder to quantify) aspects of performance, or by focusing on the quantification of outputs and outcomes at the expense of more nuanced consideration of their quality, value, and importance.

The approach presented in this guide combines theory and practice from evaluation and economics to respond to requirements for accountability and good resource allocation, as well as to support reflection, learning, and adaptive management. It involves developing and implementing a framework for:

- organising evidence of performance and VfM;
- interpreting the evidence on an agreed basis; and
- presenting a clear and robust performance story.

This guide sets out a framework for making and presenting judgements in a way that opens both the reasoning process and the evidence to scrutiny. The approach is designed to be used in alignment with broader monitoring, evaluation, and learning (MEL) systems—both for efficiency's sake, and to ensure conceptual coherence between VfM evaluation and wider MEL work.

The VfM framework achieves these aims by:

- using explicit criteria (aspects of performance) and standards (levels of performance) to provide a transparent basis for making sound judgements about performance and VfM;
- combining quantitative and qualitative forms of evidence to support a richer and more nuanced understanding than can be gained from the use of indicators alone;
- accommodating economic evaluation (where feasible and appropriate) without limiting the analysis to economic methods and metrics alone; and
- incorporating and building on an approach to VfM evaluation which is familiar to international aid donors.

[1] www.julianking.co.nz/vfi/ .

Table of contents

List of tables, figures, and boxes

List of abbreviations

ARC	African Risk Capacity
CBA	Cost–Benefit Analysis
CEA	Cost-Effectiveness Analysis
DAC	Development Assistance Committee
DFID	Department for International Development
FCDO	Foreign, Commonwealth and Development Office
GESI	Gender Equality and Social Inclusion
ICAI	Independent Commission for Aid Impact
KPI	Key Performance Indicator
MEL	Monitoring, Evaluation, and Learning
NAO	National Audit Office
OECD	Organisation for Economic Co-Operation and Development
OPM	Oxford Policy Management
ROI	Return on Investment
SNG	Sub-National Governance
TA	Technical Assistance
TMEA	TradeMark East Africa
ToC	Theory of Change
VfM	Value for Money

1 Background

This document offers practical guidance for carrying out VfM evaluations in different contexts, primarily to meet the requirements of the Foreign, Commonwealth and Development Office (FCDO), but with wider applicability to other development partners and to agencies that assess VfM in domestic policies and programmes. The approach presented in this guide combines economic and evaluative thinking to meet FCDO's requirements for VfM evaluation (DFID, 2011a; 2020) and those of other development partners.

This approach involves developing and implementing a framework for organising evidence, interpreting the evidence on an agreed basis, and telling a compelling performance story. It provides a transparent approach to making judgements from diverse evidence, which can include quantitative (economic and other indicators) and qualitative data.

Since 2016, this approach to VfM analysis has been applied by OPM to many international development projects and programmes, spanning a range of clients, countries, sectors, and budget sizes.

1.1 What is VfM?

Although there is no universal definition of VfM (King, 2019a), governmental organisations have used a range of working definitions that tend to emphasise minimising wastage, delivering outputs, achieving outcomes, improving equity, and/or maximising outcomes for a given cost—with priority often given to the latter. For example, FCDO has defined the VfM of its programme spend as making 'the best possible use of our resources to maximise our impact on poor people's lives' (DFID, 2020: 1).

More broadly, we define VfM as *good resource use* (King, 2019a). We regard VfM as an evaluative question about how well resources are used, and whether the resource use is justified (King, 2017; 2019). Evaluative questions require a judgement to be made, based on evidence and using a transparent process of reasoning. VfM cannot be 'measured' using a formula, although numeric evidence plays an important role in supporting well-informed judgements.

1.2 Why evaluate VfM?

The need to consider VfM arises because, when resources are invested in a particular policy, programme, or intervention, the opportunity to use those resources in another way is foregone. Economists call this loss of alternatives *opportunity cost* (Drummond *et al*., 2005). Consequently, choices need to be made regarding resource allocation, with a 'good' allocation being one that compares favourably to its next-best alternative (as well as meeting various other requirements such as affordability, relevance, distributive justice, human rights and dignity, and other ethical requirements).

However, the next-best alternative is not always clear or measurable, due to the complex environments in which projects operate and a lack of obvious plausible benchmarks. VfM evaluation can nonetheless determine how well the available resources are being used and whether the resource use is justified on the basis of observable features of programme delivery, short- and medium-to-long-term outcomes, and agreed definitions of what 'good' performance and value would look like—informed by comparative data, where possible.

VfM has political and bureaucratic drivers that often stem from the need for accountability and transparency in the use of public resources. Particularly in international development, because of limited aid budgets and political pressures to be accountable for the use of taxpayers' funds, donors have argued that aid should be well targeted and managed effectively. These drivers have led to an increased interest in VfM over the last 20 years (Fleming, 2013).

Good VfM evaluation can also support organisations to work effectively by helping clarify and communicate the value of their work, and by providing insights to support learning and improvement. The UK Department for International Development (DFID), which has merged with the Foreign and Commonwealth Office in 2020 to form FCDO, issued a Bilateral Review (2016b). This noted that improving VfM means not just measuring VfM and being accountable for spending and results, but also having the systems for learning, adaptation, and improvement to improve VfM at project, organisation, and system levels. The Independent Commission for Aid Impact (ICAI)'s performance review of DFID's approach to VfM (ICAI, 2018) found that the focus on VfM in UK aid and development had led to programmes finding ways to improve VfM.

Although there is increasing scrutiny on VfM in international development, significant opportunities remain to improve VfM evaluation. The objectives of this guide are therefore to assist donors and partner organisations to:

- conduct rigorous evaluations of the performance and VfM of their programmes;
- communicate findings clearly and demonstrate the value of their work; and
- support adaptive management and learning.

1.3 Context

This guide complements, and builds on, existing guides to VfM evaluation, such as FCDO's 'VfM Guidance' (DFID, 2011a; 2020), ICAI's 'Approach to Effectiveness and Value for Money' (ICAI, 2011), and Itad's 'Better Value for Money—An Organising Framework for Management and Measurement of VfM Indicators' (Barr and Christie, 2014). It also responds to ICAI's recommendations from the 2018 performance review of DFID's approach to VfM.[2]

The existing VfM guides are useful in setting out high-level principles for VfM evaluation, but fall short of providing practical guidance for assessing VfM. In particular, although both FCDO and ICAI's documents acknowledge that VfM involves making judgements from evidence, they do not set out an explicit process for reaching and presenting those judgements.

The principal contributions we aim to make in this guide are:

- to set out a framework for making and presenting judgements in a way that opens both the reasoning process and the evidence to scrutiny; and
- to facilitate the use of mixed methods evidence to support better-informed judgements, including integrating economic analysis, as well as broader quantitative and qualitative evidence, with VfM evaluations where appropriate and feasible.

[2] www.opml.co.uk/blog/value-for-money-recommendations.

1.4 Overview of this guide

Section 2 sets out the conceptual framework for the VfM approach. It highlights key concepts from evaluation and economic theories, as well as the need for the VfM approach to address FCDO's VfM criteria and to integrate this with wider MEL frameworks.

Section 3 sets out a practical approach to designing, undertaking, and reporting a VfM evaluation. It identifies a staged process involving eight discrete steps, with a particular focus on the use of predetermined criteria and standards to make judgements from the evidence.

Section 4 provides further guidance on how to develop criteria and standards.

Section 5 outlines strategies for integrating VfM evaluation with wider evaluation work.

2 Conceptual framework

2.1 Principles of the VfM approach

This section describes the key concepts that underpin the VfM approach. In brief, these are:

- **an interdisciplinary** approach, combining theory and practice from evaluation and economics;
- explicit **evaluative reasoning**, using criteria and standards ('rubrics') and providing a framework for making sound, traceable judgements about VfM;
- evidence will usually come from a **mix of methods and sources**, including existing MEL data (quantitative and qualitative) and economic analysis, if applicable, supplemented with additional evidence according to needs;
- **VfM is a matter of context and perspective**—for example, the VfM of a programme from the perspectives of recipients and in-country governments may diverge from its VfM from the donor's perspective; we therefore recommend the use of **participatory approaches** as a good evaluation practice; and
- the VfM evaluation process should be **integrated with relevant routine MEL processes and independent evaluations**, both for efficiency's sake and to ensure conceptual alignment between VfM and related MEL work.

2.1.1 Interdisciplinary approach

Traditionally, evaluation and economics have been separated by disciplinary boundaries. Both disciplines bring rich insights, but neither has all the answers when it comes to assessing VfM. Combining them can strengthen VfM evaluation. For example, evaluation contributes evaluative reasoning (Section 2.1.2), mixed methods (Section 2.1.3), participatory approaches, systems thinking and complexity-informed approaches, and a diverse toolbox for incorporating multiple values, forms of evidence, and approaches to causal inference. Economics contributes concepts such as opportunity cost, efficiency, productivity, and cost-effectiveness, as well as methods for assessing these. The starting point for our approach is to combine evaluative and economic thinking, because VfM (good resource use) is a shared domain of both disciplines (King, 2019a).

2.1.2 Evaluative reasoning

Evaluation has been defined as 'the process of determining the merit, worth or significance of something' (Scriven, 1991: 139). In plainer language, the core purpose of evaluation is to determine how good something is, and whether it is good enough (Davidson, 2005). Therefore, evaluation involves making judgements: 'it does not aim simply to describe some state of affairs but to offer a considered and reasoned judgement about that state of affairs' (Schwandt, 2015: 47).

The fundamental problem in this endeavour is 'how one can get from scientifically supported premises to evaluative conclusions' (Scriven, 1995: 51)—in other words, how to make a sound judgement using a logical and traceable process.

Explicit evaluative reasoning provides the means to make robust judgements from evidence. Essentially, it involves developing definitions of what 'good VfM' looks like on a given

programme. These definitions are developed before the evidence is gathered, providing an agreed basis among relevant stakeholders for making judgements.

Although there is more than one way to approach this task (e.g. Schwandt, 2015; King, 2019a), a widely used approach is to:

1. establish criteria of merit, worth, or significance—the aspects, qualities, or dimensions of performance that are relevant and important to forming an evaluative judgement;
2. define performance standards for each criterion to distinguish between 'excellent', 'good', 'acceptable', and 'poor' performance;
3. gather and analyse evidence of performance against the standards; and
4. synthesise the results into an overall judgement (Fournier, 1995).

The use of criteria and standards may be unfamiliar to some people conducting VfM evaluations, but exemplifies explicit evaluative reasoning, which is core to good evaluation (Yarbrough *et al*., 2011). Evaluative reasoning enhances the credibility and use of evaluation for accountability, learning, and adaptation, by providing a transparent (and therefore challengeable) basis for making judgements (King *et al*., 2013). This is the approach that is recommended in this guide, and is detailed from a practical perspective in Section 3. For further reading on evaluative reasoning, see Davidson (2014).

2.1.3 Mixed methods

A 'mixed methods' evaluation is one that combines different methods and data sources in order to arrive at a richer and more nuanced understanding than might be achieved through the use of a single method alone. For example, the Rockefeller Foundation describes mixed methods evaluation as follows (Bamberger, 2012: 1):

> *Mixed methods evaluations seek to integrate social science disciplines with predominantly quantitative and predominantly qualitative approaches to theory, data collection, data analysis and interpretation. The purpose is to strengthen the reliability of data, validity of the findings and recommendations, and to broaden and deepen our understanding of the processes through which program outcomes and impacts are achieved, and how these are affected by the context within which the program is implemented.*

Economic analysis can often provide part of the evidence needed to support a VfM evaluation (Section 5). Other numerical indicators may contribute further evidence. We have generally also found, once criteria and standards are established, that some of the evidence needed to support evaluative judgements is usually qualitative. Indicator-based (quantitative) measurement makes a valuable contribution to evaluating programme performance and VfM. Indicators alone, however, are often insufficient to support well-reasoned evaluative judgements about a complex programme. Indicators, by their very nature, are narrow and provide individual pieces of measurable evidence that may correlate with some criteria. Broader contextual (qualitative) evidence is also important to provide further information about performance and support appropriate interpretation of the indicators.

To support sound VfM evaluation, we need to draw on a balance between quantitative measures and qualitative data. A mix of evidence can give us a better understanding of whether a programme is really changing lives.

Five key benefits of using mixed method designs (Greene, 2005) include:

- **triangulation of evaluation findings**: enhancing the validity and credibility of evaluation findings by comparing information obtained from different methods of data collection;
- **development**: using the results of one method to help develop the sample or instrumentation for another;
- **complementarity**: extending the comprehensiveness of evaluation findings through results from different methods that broaden and deepen the understanding reached;
- **initiation**: generating new insights into evaluation findings through results from different methods that diverge and thus call for reconciliation through further analysis, reframing, or a shift in perspective; and
- **value diversity**: incorporating a wider diversity of values through the use of different methods that themselves advance different values.

2.1.4 Participatory approaches and perspectives

> *Beneficiary feedback has become part of the fabric of good programme delivery.* (DFID, 2016b: 46)

VfM has traditionally been assessed from the donor's perspective, with a focus on donor resources spent and the achievement of outputs and outcomes specified by the donor. However, other parties also invest resources (monetary and non-monetary) in programmes, and the value derived from the investment (including what is considered valuable, and why) is a matter of context and perspective. Moreover, 'in a social investment, those who invest and those who benefit are different people and may have very different perspectives about what constitutes a return on the investment' (King, 2019a: 32).

For example, an FCDO-funded programme may use health workers funded by the country government to support implementation. The cost of these health workers needs to be captured in the total cost of the programme. For the local community, the opportunity cost of these health workers may be very high if it diverts some of their time from other important healthcare work.

In order to provide good VfM evaluations that help to make limited aid resources count and support investment in approaches that are effective, we need to consider how power is shared, which values matter, and how to use those values systematically to support good decisions that have positive impacts for communities. Like any good evaluation, VfM evaluation should seek to empower those who are supposed to benefit from an intervention—in determining criteria, gathering evidence of performance, and using the criteria and evidence to evaluate performance and VfM.

While it will not always be feasible to consider VfM from multiple perspectives, we believe the need to do so should be carefully considered during the VfM design process, to determine whether the VfM evaluation design (including criteria, standards, and methods for gathering evidence) 'is justified in the cultures and contexts where the programme has consequences' (Yarbrough *et al*., 2011: 251).

2.2 Integration with MEL systems and evaluations

Our approach views VfM as an evaluative question, and recommends evaluative methods to guide the evaluation. In this context, VfM is an extension of evaluation that adds the dimension of resource use to some measure or description of benefit, to make a judgement about whether resources are being used in a worthwhile way.

VfM assessment is often conducted as a standalone exercise, and throughout this guide we discuss VfM design and implementation largely in this context. However, VfM may also be part of an MEL cycle, with recurrent evaluations conducted in the project lifecycle. Alternatively, it may be part of less frequent (or one-off) formative or summative evaluations. When VfM evaluations are part of broader evaluations or MEL cycles, careful thought is needed to integrate VfM at the outset to avoid duplication wherever possible. This is important for two reasons.

- Conceptually, the VfM and evaluation/MEL frameworks should be coherent with each other; for example, they should be based on the same Theory of Change (ToC) and should use criteria and indicators that are consistent (or at least not contradictory).
- Practically, the VfM and evaluation/MEL frameworks and reporting should be coordinated to avoid overlap and duplication of effort, and to ensure the right information is collected at the right time to serve both purposes.

Accordingly, VfM and MEL frameworks should be developed through a single, joint design process to determine boundaries and promote coherence—e.g. to identify *all* evaluation objectives and questions (agree boundaries between them; develop a shared ToC; and plan for a coordinated reporting approach).

Given the diverse demands of evaluation clients and the plurality of evaluation approaches, **the VfM approach needs to be adaptable so it can be integrated with the evaluation context of each programme**. We discuss integrating VfM with evaluation and MEL systems in more detail in Section 5.

2.3 Timing/frequency of VfM evaluations

When this VfM approach is used to support regular VfM evaluations and to monitor and track VfM over time, the frequency should be determined by the nature of a programme's intended outcomes and the expected timeframes to achieve substantive outcomes. For example, it might make sense to track some outcomes on an annual basis, whereas for impacts that take longer to achieve, assessments carried out every two to three years may be more realistic if shifts are to be detected.

Assessments of economy and efficiency may be carried out more frequently (e.g. six-monthly) for diagnostic purposes if desired.

2.4 Level of effort required

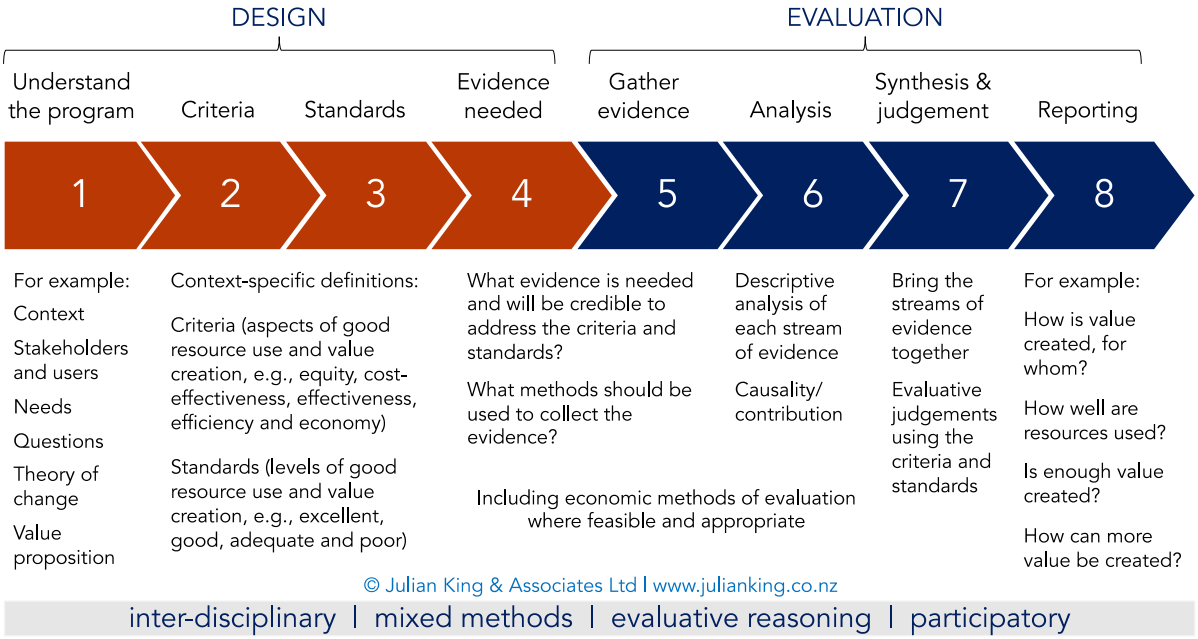
A full VfM evaluation using this framework can be a time-intensive undertaking, and consideration must be given to proportionality when assessing the VfM of smaller programmes. The use of criteria and standards to guide evaluative judgements must be tailored to the scope and scale of the programmes, as well as to available resources and timeframes for conducting VfM evaluations. The reasoning process remains the same, while the scope of the criteria and the content and comprehensiveness of the evidence can be varied, remaining transparent and open to ongoing revision.

3 Practical approach

This section sets out a practical approach for designing, undertaking, and reporting a VfM evaluation. It identifies a staged process involving eight discrete steps, with a particular focus on the use of predetermined criteria and standards to make judgements from the evidence. Throughout the process, opportunities to align VfM framework design and reporting with wider MEL activity are identified.

The key steps involved in explicit evaluative reasoning are summarised in Figure 1 and explained below. Note that Steps 1–4 relate to the design of the VfM framework, while Steps 5–8 relate to VfM evaluation and reporting, which can only be done once the framework is agreed and in place.

Figure 1: Overview of the VfM approach[3]



An evaluation framework of this type is designed to address one or more key evaluation questions—high-level, big picture questions that reflect the purpose of the evaluation. In a VfM evaluation, the key evaluation question would take the following general form: *To what extent does the programme represent value for the resources used, and how can its value be improved?*

The general stages involved in VfM framework design include developing an understanding of the programme (including specifying or reviewing a ToC and a value proposition), identifying VfM criteria, developing standards for each criterion, and determining the evidence necessary (and available) to support well-informed evaluative judgements against the criteria and standards.

[3] Adapted from www.julianking.co.nz/wp-content/uploads/2014/09/Screen-Shot-2016-11-27-at-4.36.04-PM.p.

3.1 Step 1: Understand the programme

At the commencement of any evaluation, the evaluation team should invest time in understanding the programme, its context, and evaluation users and stakeholders, as well as their information needs. Understanding the programme usually includes reviewing or developing its ToC to ensure the evaluation starts with a clear and shared view of what needs the programme is intended to meet and how it is intended to function.

Theory of Change

A ToC 'explains how activities are understood to produce a series of results that contribute to achieving the final intended impacts' (Rogers, 2014: 1). One of the functions of a ToC is to assist in the identification of criteria, standards, and indicators that are relevant to the programme's results chain.

By following the sequence shown in Steps 1–4 (starting with a ToC before moving on to the development of criteria and standards and the selection of metrics and methods), we can ensure that the values embedded in the criteria and standards reflect the programme theory and that the metrics in turn reflect the content of the criteria and standards. This helps ensure that the indicators we use are valid, cohere with the programme theory, and provide relevant evidence to support contribution analysis and evaluative judgements.

This guide does not cover how to develop a ToC, as this is well covered in other texts (e.g. Rogers, 2014). Both the VfM evaluation and any other MEL activity should use the same ToC so that the evaluation frameworks are consistent and coherent across all MEL activity. This may affect decisions about when to develop the ToC and who should be involved (e.g. MEL staff as well as the person who will be in charge of VfM evaluation, together with any other stakeholders who would normally be involved).

Typically, a ToC will be developed during the inception phase of a project, so it makes sense to integrate the development of the VfM framework with this activity. However, if a VfM framework is being developed once the project has already begun implementation, it will nonetheless be necessary to review the ToC to ascertain whether it is fit for purpose, adding to or amending it as necessary.

Box 1: Example: Reviewing a ToC for a VfM framework

For example, the Pakistan Sub-National Governance (SNG) Programme supported local governments in two provinces in Pakistan (Khyber Pakhtunkhwa and Punjab) to improve the delivery of basic services. Financed by FCDO and managed by OPM, the programme operated in 12 districts (six each in Khyber Pakhtunkhwa and Punjab), supporting reforms in public financial management, governance, and planning, and operating a challenge fund to finance innovative service improvement pilots. The intended outcome of the programme was that 'poor people in Punjab and Khyber Pakhtunkhwa report that services are better meeting their needs', and its intended impact was 'a more stable democracy in Pakistan, through increased trust in Government'.

As part of the SNG Programme, the ToC set out in the DFID business case defined outcomes at a high level ('services better meet the needs of poor people in Punjab and Khyber Pakhtunkhwa'), leaving a relatively long conceptual journey from outputs to the intended outcome. The VfM consultants worked with SNG MEL advisers, technical advisers, and management to review the ToC at intermediary outcome level, unpacking the implied causal pathway towards outcomes and validating the intermediary outcomes with the project teams.

It was also noted that the value of the SNG Programme did not only lie in the achievement of the specified outcomes. Critically, it also involved its capacity to generate learning that influenced other relevant programmes, as well as its capacity to be adaptive, responding to lessons learnt as well as

emergent opportunities and challenges as they arose, maximising the value of the DFID-funded technical assistance (TA). These features were therefore also included in the VfM framework (King and Allan, 2018).

Incorporating value creation in a ToC

A good ToC is explicit about resources and inputs and how they are supposed to be used to produce outputs and outcomes, so may already contain a lot of the information needed to support the identification of VfM criteria and standards. However, outcomes and value are distinct (though sometimes interrelated) concepts. For example, a small impact can create a lot of value, or *vice versa*. When designing a VfM framework, we have found that extending the ToC to consider how value is created (as well as what kinds of value, for whom) can enhance clarity.

A theory of value creation, or value proposition, seeks to 'describe the mechanisms by which new, transformed or superior value is created by the resources consumed by a program. This extra theoretical layer represents an advance over conceptualizing programs as vehicles for "making a difference" to programs as transformational processes that convert resources (funding, expertise, relationships, etc.) into significant social value' (King, 2021: 6).

Incorporating value creation in a ToC does not need to be complicated. It can be as simple as providing concise narrative answers to the following questions.

- What is the programme's value proposition (a value proposition is a declaration of intent that communicates how people should benefit from the programme)?
- What kinds of resource are invested in the programme, and by whom?
- What kinds of value does/should the programme create, and from whose perspective does this constitute value?
- How is value created? What are the mechanisms by which the programme uses resources efficiently, effectively, and equitably, and creates sufficient value to justify the investment?
- How is value distributed?
- What factors influence the extent to which resources are transformed into worthwhile value?

A note about 'different kinds of value'. Value (also described as merit, worth, and significance) is distinct from impact because processes and impacts do not have intrinsic value: people place value on them (Gargani, 2018). Accordingly, value is a matter of context (where, when) and perspective (to whom). Various different taxonomies of value have been proposed, but we can intuit that programmes and their impacts matter to people in different ways. For example, a programme can create social, cultural, spiritual, environmental, or economic value. These categories are not exhaustive: they can overlap, and there are subcategories within them. For example, social value could be related to social cohesion, freedom, inclusion, justice, etc.

Box 2: Example: African Risk Capacity (ARC)'s value proposition

In the evaluation of the ARC programme, we developed and tested ARC's value proposition with stakeholders, informed by past evaluations, literature, formal theory, and explicit assumptions. Articulating a shared understanding of the programme's value proposition helped with co-developing contextually appropriate VfM criteria and standards articulating what good VfM looks like. This in turn aided in the selection of an appropriate mix of methods (qualitative, quantitative, and economic) to gather and analyse credible evidence. Having an explicit value proposition provided additional guidance for making valid VfM evaluative judgements that could get to the heart of what it means for ARC to create value.[4]

The following critical, observable factors affecting the potential VfM of ARC's drought insurance were identified from the value proposition:

- economy: capital, reinsurance, and running costs are kept as low as possible, and ARC offers relevant, well-priced products;
- efficiency: accurate prediction of relevant drought events; reliability and timeliness of ARC payouts; ARC provides high-quality TA and capacity building support to countries to improve drought preparedness;
- effectiveness: ARC support to enhanced contingency planning and capacity building contributes to timely and appropriate drought response; ARC-funded drought response reduces negative household-level coping strategies; ARC successfully facilitates increased sovereignty and sustainability in country-level risk planning; and
- equity: ARC-funded drought response support reaches the people who are most in need.

3.2 Steps 2 and 3: Develop VfM criteria and standards

As noted in the previous subsection, evaluation requires not only sound evidence, but also an explicit basis for making evaluative judgements. Criteria and standards (organised into tables called 'rubrics') provide an agreed, transparent basis for interpreting the evidence and making these judgements. In this context:

- **criteria of merit or worth** are selected dimensions of performance that are relevant to a particular programme and context. They describe, at a broad level, the aspects of performance that need to be evidenced to support an evaluative judgement about VfM. When using the 5Es as the basis for VfM evaluation (see Section 4.2), context-specific definitions of economy, efficiency, effectiveness, equity, and cost-effectiveness need to be developed; and
- **performance standards** provide defined levels of VfM for each of the criteria. In our VfM evaluations, we typically label these levels as excellent, good, adequate, and poor. However, in some cases, an alternative framing (e.g. stages of growth such as emerging, evolving, embedding, and excelling) may be preferred.

Criteria and standards should reflect the key inputs, activities, deliverables, outputs, intermediary and long-term outcomes, impact, and value articulated in the ToC and value proposition. They should be agreed with key stakeholders in advance of the VfM evaluation.

Section 4 provides more detailed guidance on selecting and defining VfM criteria and standards. Programme-specific definitions of all criteria need to be developed, together with standards describing different levels of performance for each criterion. However, the 5Es

[4] www.opml.co.uk/projects/independent-evaluation-african-risk-capacity.

often provide a reasonable starting point for developing contextually appropriate VfM criteria that will meet the requirements of FCDO and many other donors.

Developing VfM criteria and standards around the 5Es involves the following steps.

1. For each of the overarching criteria (economy, efficiency, effectiveness, equity, and cost-effectiveness), write a concise programme-specific definition.
2. For each definition, identify subcriteria that describe dimensions of good performance and VfM.
3. Organise the subcriteria into levels (e.g. definitions of excellent, good, adequate, and poor) in a table.

This approach is flexible and needs to be tailored to context. While this flexibility is a key strength of the approach, it also makes it difficult to generalise about how rubrics should be developed and what they should look like. The following tips, therefore, are guidelines, not rules.

Use participatory processes

The process of developing criteria and standards can be participatory and/or draw from programme documentation and literature. The best approach, where feasible, is to combine these elements. Running a participatory process requires skilled facilitation. During the design of an evaluation, stakeholders can be brought together to identify the values that form the basis of the criteria and standards. Although key requirements will be formally documented (e.g. in a business case, proposal, contract, logframe, or ToC), the criteria and standards will elaborate on these requirements, which may involve reconciling or accommodating diverse perspectives (King *et al*., 2013).

A shared understanding of what matters is reached through this process, and this is articulated in the criteria and standards. This stage is iterative and may involve progressing through several drafts. Taking the time to reach this agreement is 'an early investment that pays dividends throughout the remainder of the evaluation' (King *et al*., 2013: 14). Through their participation, donors and stakeholders become engaged in the evaluation design, see their expectations represented in the criteria and standards, and understand the basis on which judgements will be made (King *et al*., 2013).

This approach to evaluation therefore has a number of advantages beyond supporting robust reasoning (King *et al*., 2013), as follows.

- When the basis for judging performance and VfM is articulated and agreed at the start, it increases the likelihood that findings will be endorsed by donors and stakeholders.
- If there are diverse viewpoints and values, it is desirable for donors, delivery teams, and communities to have transparent discussions about these early on, before the evaluation is conducted, with a view to reaching a shared understanding of what matters—or at least to acknowledge and accommodate different perspectives.
- When donors and other stakeholders are involved in the process of making evaluative judgements using the criteria and standards, it increases transparency about how the judgements are made.
- The criteria and standards provide a basis for structuring a concise and focused report.
- Donors find this type of evaluation easy to use and credible.

Get the right people in the room

When using participatory processes, it is important to consider who should be at the table. Jane Davidson (2014:8) offers the following guidance:

Who should be involved in the evaluative reasoning process itself? Key considerations include the following.

- **Validity**: Whose expertise is needed to get the evaluative reasoning right? As well as a lead evaluator with expertise in evaluation-specific methodology, it will be necessary to include people with expertise in the subject matter; in the local context and culture; and in interpreting the particular kinds of evidence to be used.
- **Credibility**: Who must be involved in the evaluative reasoning to ensure that the findings are believable in the eyes of others?
- **Utility**: Who is most likely to use the evaluation findings (i.e. the products of the evaluative synthesis)? It may be helpful to have primary intended users involved so that they can be part of the analysis and see the results with their own eyes. This increases both understanding of the findings and commitment to using them.
- **Voice**: Who has the right to be at the table when the evaluative reasoning is done? This is particularly important when working with groups of people who have historically been excluded from the evaluation table, such as indigenous peoples. Should programme/policy recipients or community representatives be included? What about funders and programme managers/frontline staff?
- **Cost**: Taking into consideration the opportunity costs of taking staff away from their work to be involved in the evaluation, at what stages of the evaluative reasoning process is it best to involve different stakeholders for the greatest mutual benefit?

In practice, some of these considerations may be more important for a full summative evaluation of a programme than for a periodic VfM evaluation. However, in principle, these points are worth considering for every evaluation, including VfM.

The realities of designing a VfM evaluation within time and budget limitations mean that an inclusive, iterative process will not always be possible. At a minimum, time should be taken to ensure the donor and other key stakeholders fully understand and endorse the approach and the basis for making judgements, even if they are not present during the development of the criteria and standards. A VfM report is more likely to be accepted and considered valid if all the evidence and reasoning is traceable, open to scrutiny, and linked to criteria that were developed (or at least discussed) with stakeholders at the start.

Tailor to context

Good criteria are specifically tailored to the programme—and generally at a specified point in time (e.g. what would 'good' look like at the time of the VfM evaluation?). This means that the specific definitions of economy, efficiency, etc., will be different for each programme.

The temptation to copy and paste definitions from other VfM frameworks should therefore be avoided. That being said, some concepts are likely to be transferrable, so it is worth scanning past VfM frameworks for inspiration.

Keep it simple

Criteria and standards need to be specific enough to guide meaningful judgements, yet not over-specified to the point that they make us focus on the wrong things. For example, we must not make unwarranted assumptions about programme results that are not knowable in advance.

Furthermore, criteria and standards need to be kept manageable so that VfM evaluations do not consume time and resources beyond what can be considered 'commensurate' with the overall programme size. Instead of describing every possible dimension of a programme's intended performance, it is better to think in terms of the '80/20 rule': what are the top 20% of possible criteria that would represent 80% of the important features of the programme's performance and VfM? This may require careful balancing and prioritisation of diverse stakeholder interests and priorities, avoiding the temptation to include all suggestions from all stakeholders. Careful planning and costing of evaluations against the resources available can also help keep the overall scope, and the number of criteria/subcriteria, manageable.

Aim for consistency in performance levels

Although criteria and standards are tailored to a specific programme and context, we should strive for a consistent interpretation of what 'excellent', 'good', 'adequate', and 'poor' performance means in every VfM evaluation, as far as possible. See Table 2, Section 4.4 for guidance.

Suspend conversations about measurement

During criteria development, it is important to 'park' any conversations about evidence, such as which indicators may be used and how they will be measured. This question is dealt with later on.

The purpose of rubrics is to describe what aspects of performance or VfM are important (criteria) and what they would look like at different levels of performance (standards). These descriptions differ from measurable indicators and other forms of evidence in that they focus on the intended functioning and effects of the programme, rather than on how they might be measured. While indicators are specific and measurable, criteria and standards describe the nature of the actual changes that are intended, which are generally described in broader, less specific terms in order to facilitate meaningful evaluative judgements:

> *Good rubrics … encourage the use of sound evaluative judgement, providing a shared language for interpreting evidence that increases both inter-rater reliability and validity. This approach contrasts with many approaches to 'quantification' that try to eliminate judgement as if it is inherently unreliable and based solely on opinion.* (Davidson, 2014: 11)

Reference existing benchmarks (where appropriate)

In many sectors, generally accepted benchmarks exist that define sector-specific criteria and standards, such as the World Bank's International Benchmarking Network for Water and Sanitation Utilities; UNICEF's Child Well-Being Indicators; the United States Agency for International Development's Health Systems Benchmarking Tool; and the Public Expenditure and Financial Accountability Framework for public financial management. These existing benchmarks should be consulted and referenced where appropriate so that

the definitions of 'poor', 'adequate', 'good', and 'excellent' performance are aligned with relevant points of comparison.

Rubrics are living documents

Rubrics do not have to be treated as 'finalised' and can be transparently amended during the course of the evaluation. As with a ToC, rubrics tend to be developed during evaluation design. Later, as the programme and the evaluation evolve, new information may emerge that warrants revisions to the ToC and/or to the rubrics. To guard against perceptions of 'moving the goalposts', it is important to document any updates to rubrics transparently, with supporting rationale.

3.3 Step 4: Identify the evidence required and select methods

In a logical and sequential process of evaluation design, it is only after clarifying the ToC, criteria and standards that relevant sources of evidence can be identified. The preceding steps are important to help ensure that the evidence is relevant, measures the right changes, and is appropriately nuanced.

This step involves determining what evidence is needed and will be credible to address the criteria and standards, and what methods should be used to gather the evidence. This involves systematic analysis of the criteria and generally involves asking questions such as: How will we know? What would be credible evidence? What evidence is available or feasible to collect (King *et al*., 2013)?

What counts as credible evidence to tell us whether we are seeing poor, adequate, good, or excellent VfM performance in the programme? This requires consideration, not just of things we can count or data we happen to have available to us, but also of what really matters and has relevance to the VfM of the programme. Ideally, this will include a mix of numeric, qualitative, economic, and monetary evidence.

Key sources of evidence to support the VfM evaluation will vary with context but may, for example, include programme financial accounting data, logframe indicator data, programme operational data/reports, MEL data/reports, and results of economic analysis.

From a pragmatic perspective, there should be a preference for using data that are already produced, where they are appropriate. For example, there may be existing reporting mechanisms against MEL frameworks and contractual key performance indicators (KPIs). Accordingly, it can be helpful under 'Step 1: Understand the programme' also to assess existing evidence sources and reporting requirements. This is discussed further in Section 5.2.

Prepare a VfM framework

Steps 1–4 collectively outline the key tasks involved in designing a VfM framework. At the end of Step 4, the VfM framework can be written up. As with any evaluation framework, the VfM framework should describe:

- the programme (including its ToC/value creation in an annex);
- the criteria and standards that will be used to assess the programme's VfM;
- the evidence sources that will be needed and the methods that will be used to gather the evidence (including any linkages to the MEL framework); and

- a plan and timetable/cycle for conducting the VfM evaluation(s).

3.4 Step 5: Gather evidence

Gathering the evidence needed for a VfM evaluation involves following the same principles of good project management and fieldwork that would be followed for any evaluation or research project. If the VfM framework is well aligned with wider MEL frameworks, the VfM evaluation may be able to make use of relevant MEL data, minimising duplication of effort.

Causality or contribution

As with any evaluation, it is important to consider the counterfactual when making judgements about the attribution of results to a programme—i.e. what would have happened without the programme? Some programmes will involve experimental (e.g. randomised controlled trial) or quasi-experimental evidence to support causal inferences. In other cases, it will be necessary to apply a theory-based approach, such as contribution analysis (Mayne, 2008), or other frameworks to guide well-reasoned causal inferences such as the Bradford Hill Criteria[5] (strength, consistency, specificity, temporality, dose-response gradient, plausibility, coherence, experiment and analogy) or Davidson's 8 strategies, which are consistent with Bradford Hill and also include asking observers, controlling statistically for extraneous variables, and looking for "telltale patterns" (Davidson, 2005).

It is beyond the scope of this guide to provide a full menu of options to assess causality/contribution. At a minimum, VfM evaluation should seek to provide sufficient evidence and reasoning to convince a reasonable but sceptical observer whether the programme has contributed to the observed results (Donor Committee for Enterprise Development, 2016).

Using 'additionality' concepts can help to guard against over-claiming a programme's impact. Considerations include:

- Deadweight – to what extent might observed outcomes have occurred without any intervention?
- Shared effects/contribution – to what extent might other interventions or programmes have influenced changes?
- Gains through positive externalities – did the programme contribute to verifiable indirect benefits, or generate positive benefits for other parties?
- Losses through displacement, substitution, leakage, or negative externalities – did changes occur that cannot be claimed as programme benefits, or did the programme generate negative effects or costs for other parties?
- Sustainability – are results expected to increase, stay the same, or drop off over time?

[5] For example, see: https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095523346

An example of this checklist in use (which was adapted from SROI Network, 2012 and UK Government, 2014), can be seen in OPM's VfM assessment of TradeMark East Africa.[6]

In evaluations of less complex programmes it may make sense to assign a percentage contribution to observed results. However, this procedure should be approached with caution, as it is not always meaningful to do so, especially in situations where different programmes may act synergistically. It is more important to be able to explain credibly *how* the programme contributes to the results seen.

3.5 Steps 6–7: Analysis, synthesis, and judgements

These stages involve three discrete, sequential steps.

- **Analysis** of each stream of evidence (quantitative and qualitative) is first carried out to identify individual findings and themes.
- **Synthesis** is then undertaken to triangulate and consider the totality of evidence collected, including any areas of corroboration or contradiction between evidence sources. This is first done for each VfM criterion individually (i.e. economy, efficiency, effectiveness, cost-effectiveness, and equity), and second for VfM overall.
- **Judgements** are then made against the criteria and standards (i.e. the rubric) to determine and transparently report the VfM of the programme.

Critically, the overall judgement of VfM is not required to be a simple average of the judgement against each criterion. Rather, greater weight should be given to those criteria that are deemed more relevant, as agreed at the outset by programme stakeholders according to what is most valued on the specific programme.[7] This may also reflect the timing of the VfM evaluation. For example, when evidence on effectiveness and cost-effectiveness is available later in the programme lifecycle, these criteria would generally be given greater weight than evidence focusing on inputs (e.g. economy) and outputs (e.g. efficiency), which receive more weight at early stages of implementation. Conversely, in the early stages of a programme, economy and efficiency may be the only criteria included in the VfM evaluation.

If sufficient upfront investment is made in developing sound criteria and standards, it is in the synthesis of findings that the investment pays off, because criteria and standards provide a framework for systematically interpreting the multiple streams of evidence. As the criteria were predetermined and the streams of evidence were deliberately identified to relate to the criteria, the streams of analysis all map back to the criteria in a logical way (King *et al*., 2013). This makes it relatively straightforward to make judgements against the individual criteria, and then for VfM overall.

In a rapid VfM evaluation of a small programme, the sensemaking and reporting steps need not be onerous. For example, the VfM evaluation may involve a two-hour workshop with key programme staff to reflect on existing evidence of programme performance, understand the context behind findings, and make judgements against the criteria and standards with supporting rationale.

[6] https://www.trademarkafrica.com/wp-content/uploads/dlm_uploads/2020/01/TMEA-Value-for-Money-Assessment-FINAL.pdf

[7] See, for example, OPM's VfM assessment of TradeMark East Africa cited above.

Who should make the judgements?

Judgements can be made by a programme's management or MEL team, by the donor, by independent evaluators, or, preferably, by some combination of these. As the initial judgements (and the criteria, standards, and evidence upon which they are based) are open to scrutiny, they can be validated, contextualised, and challenged by funders and other stakeholders.

Programme MEL, donor, and independent evaluator perspectives are all important and will add value to a VfM evaluation. As noted earlier, it would be highly desirable to include a beneficiary perspective in many circumstances. It is important to have an independent view on these judgements, but supported by the in-depth contextual and evidence base from those directly involved in the programme. A preferred forum, therefore, is a joint meeting or workshop facilitated by independent evaluators.

The judgement-making process can be viewed as an opportunity for the programme to engage with donors in reviewing the evidence, and in judging performance and VfM against the agreed criteria and standards. The general approach in this type of workshop is to present the evidence (the 'what's so') and an initial synthesis to participants and facilitate a process of collective sensemaking. The purpose of this process is to reach a shared understanding about what the findings mean and the level of VfM to which the evidence points (the 'so what?') and to reflect on any actions that should be taken to improve performance and VfM (the 'now what?'). Throughout this process, the criteria and standards provide a focal point and a framework for systematically considering the evidence and making judgements.

A few notes on making judgements

The process of making judgements can feel a little unfamiliar to those who have not used rubrics before, because it involves using evidence in a new way: comparing the evidence to the criteria and standards, and being deliberative about what level of performance it points to. An evaluative judgement cannot be made by an algorithm as it involves weighing multiple pieces of evidence—some of which may be ambiguous or contradictory—guided by the criteria and standards, with the aim of making transparent and defensible judgements with a clear rationale.

In some cases, different streams of evidence may point to aspects of performance at more than one level of a rubric. This is normal. The following pattern-spotting questions may assist in making evaluative judgements (Eoyang and Oakden, 2016).

- In general, to what level of performance does the evidence point?
- Are there any outliers or exceptions that modify your judgement?
- Are there any significant contradictions to weigh up (on the one hand, on the other hand)?
- Check your own assumptions/bias—does anything surprise you?
- Is anything puzzling? Is more evidence needed?

The key question to ask is where the centre of gravity sits overall. If in doubt, choose the lower of two levels. You can always include a qualifying statement—for example, 'the evidence indicates that the programme meets nearly all of the criteria for excellent efficiency, but is held back by one important issue that needs to be addressed, and therefore a judgement of good efficiency has been reached.'

Perceived 'subjectivity' of judgements

The use of criteria and standards to make judgements about performance from the evidence may be perceived by some stakeholders as being 'too subjective' or 'less robust' than a purely measurement-based system. This is incorrect. Two separate issues need to be understood: evaluative reasoning, and evaluation methods.

- *Evaluative reasoning is the means by which judgements are made*: The purpose of rubrics is to set out an explicit and agreed set of criteria and standards, defined in advance of the VfM evaluation. This guards *against* personal subjectivity in judgements because the basis for interpreting the evidence is transparent. It would be more correct to describe a rubric as *inter-subjective*—that is, an agreed construct used by a group of people for an agreed purpose (King, 2023a).[8]
- *Evaluation methods are the means by which evidence is gathered*: Our approach to VfM recognises the validity of diverse methods and forms of evidence, qualitative and quantitative, and the use of multiple evidence sources for triangulation. It would be incorrect to assert that quantitative data is inherently objective and that qualitative data is subjective. Either form of evidence can be used objectively or subjectively. For example, the selection of indicators in logframes can be highly subjective, as can the interpretation of the indicator data.

The use of rubrics supports clear evaluative reasoning *and* sound use of mixed methods evidence.

3.6 Step 8: Reporting

A good VfM report:

- tells a compelling performance story focused on, and structured around, the aspects of performance that matter (as defined by the criteria), and presenting a clear judgement about the level of performance (as defined by the standards); and
- gives clear answers to important questions by getting straight to the point, presenting transparent evidence, and being transparent about the basis upon which judgements are made (McKegg *et al*., 2017).

Accordingly, VfM reports should be structured around the overarching VfM criteria (e.g. the 5Es), addressing each criterion systematically in turn.

The first page of the report should present a summary of findings (e.g. a dashboard) and the body of the report should present each criterion sequentially, setting out a definition, relevant criteria and standards, a summary of evidence against them, and a judgement on performance.

Formative VfM evaluation: identifying opportunities to improve

The process of making judgements about VfM against criteria and standards provides a way of systematically identifying areas where a programme is performing well and, potentially, areas where improvements can be made to strengthen VfM. To illustrate, a programme might meet all but one of the criteria for 'excellent' efficiency. If so, the lagging criterion might

[8] https://en.m.wikipedia.org/wiki/Intersubjectivity.

pull the judgement down to 'good' or 'adequate', and would be identified as an area for improvement. The VfM report should identify these opportunities.

4 Guidance on developing VfM criteria and standards

As outlined in preceding sections, our approach hinges on developing agreed definitions of good performance and VfM, including criteria (aspects of performance) and standards (levels of performance). Criteria and standards, aligned with the design and context of the policy or programme, provide a framework for organising evidence of performance and VfM, interpreting the evidence on an agreed basis, and presenting a clear and robust performance story.

The previous section included principles for developing criteria and standards: use participatory processes; get the right people in the room; tailor to context; keep it simple; aim for consistency in performance levels; suspend conversations about measurement; reference existing benchmarks (where appropriate); and treat rubrics as living documents.

This section focuses on the content of VfM criteria and standards. It outlines considerations that can guide the development of contextually appropriate criteria and standards, and gives examples of criteria and standards from past VfM evaluations.

4.1 VfM criteria

VfM criteria are selected dimensions of programme performance that are relevant to good resource use. They describe, at a broad level, the aspects of performance that need to be evidenced to support an evaluative judgement about VfM.

As with any evaluation, VfM criteria need to be contextually determined. They need to reflect dimensions of performance of the project, programme, or policy being assessed that contribute to good VfM (good resource use).

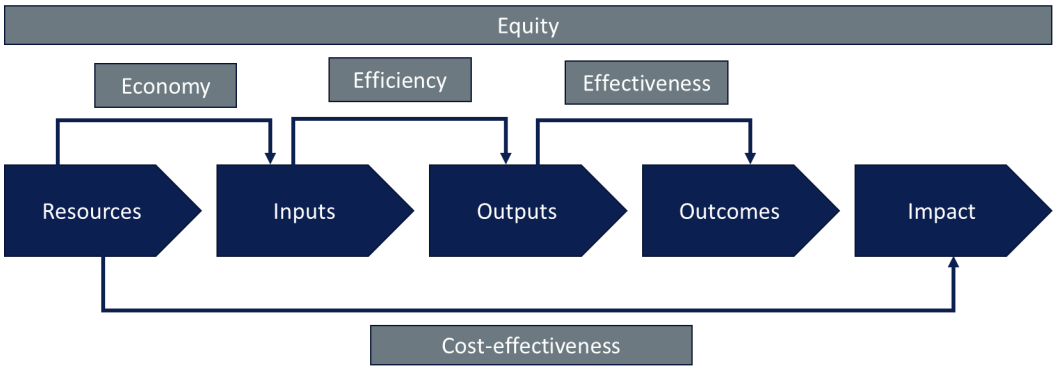
For a comprehensive VfM evaluation, the selection of criteria should help us respond to questions such as:

1. How well have resources been used?
2. What value has been created?
3. Does the value created justify the resource use?
4. How can the resource use be improved?

In addition to making summative judgements, we should determine whether there are opportunities to improve resource use (the fourth question in the above list), whether through making changes to an intervention or proposing an alternative intervention altogether. As detailed earlier, the purpose of assessing VfM is to find the most effective way of using resources to deliver intended outputs and outcomes. Therefore, the aspects of performance covered by the criteria should reflect areas that are likely to generate meaningful learning and inform decisions—that is, VfM criteria and evaluations should be utilisation-focused (Patton and Campbell-Patton, 2021).

4.2 5Es as a starting point

Figure 2: Adapted from the FCDO VfM 5Es conceptual framework



There are no prescribed criteria that must be used in the evaluation of VfM. However, FCDO and UK National Audit Office (NAO) guidance recommend the use of the 5Es: economy, efficiency, effectiveness, equity, and cost-effectiveness. While it may occasionally be appropriate to consider additional criteria or to only use some of the 5Es, the framework provides a useful starting point for developing VfM criteria.

Table 1: Definition of the 5Es (DFID, 2020)

Criterion	Definition
Economy	Are we (or our agents) buying inputs of the appropriate quality at the right price?
Efficiency	How well are we (or our agents) converting inputs into outputs?
Effectiveness	How well are the outputs produced by an intervention having the intended effect?
Equity	How fairly are the benefits distributed? To what extent are we reaching marginalised groups?
Cost-effectiveness	What is the intervention's ultimate impact on poverty reduction, relative to the inputs that we or our agents invest in it?

The 5Es framework is relevant at all stages in the programme life cycle. In design, the business case focuses on potential cost-effectiveness and efficiency. The 'design value' of a programme is set in the appraisal phase, and procurement emphasises the selection of service providers who can deliver economically and efficiently. Programme implementation aims to balance all 5Es, though it is reasonable to be more reliant on process indicators than on results to assess VfM in the early stages of a programme.

The inclusion of equity in the framework is particularly important. The core objectives of aid programmes commonly include addressing inequities (between countries and/or groups within a country). Therefore, the VfM of development aid does not just depend on efficient delivery and cost-effective outcomes, but also on improving distributive justice and upholding human rights and dignity among different groups, especially the most marginalised (which varies by context and may include women and girls, young people, poor people, people with disabilities, and people living in more remote rural areas). In a great many cases, VfM is fundamentally concerned with delivering improvements in equity as efficiently and cost-effectively as possible. Achieving this involves a balancing act. For example, working to improve the lives of the most disadvantaged may be costlier per beneficiary than improving

the lives of moderately disadvantaged people. This underscores the need for evaluative judgements about VfM, balancing multiple criteria and recognising trade-offs between them.

4.2.1 Limitations of the 5Es

The Three Es of economy, efficiency, and effectiveness divide the concept of VfM into discrete sections of a programme's results chain. Economy focuses on the cost and value of inputs. Efficiency focuses on the transformation of inputs, by sets of activities, into outputs. Effectiveness focuses on the achievement of outcomes. Cost-effectiveness looks at the whole results chain from inputs through to impact. While this approach offers a pragmatic breakdown of VfM criteria, a simplified breakdown of inputs, outputs, and outcomes may not adequately capture the realities of aid programmes where change is often non-linear, where adaptive management involves trialling multiple strategies in order to find a few that work, or where some of the value created is intangible (e.g. the value of programme learning).

The simplified results chain also omits some steps that may be important for VfM. For example, in between inputs and outputs, a full ToC would specify programme activities and deliverables that would help to explain how the programme is supposed to deliver results and VfM. Similarly, the journey between outputs and outcomes may be conceptually quite long, and it may be desirable to specify one or more levels of intermediate outcomes that sit between the programme's outputs and long-term outcomes.

Critically, economy is neither necessary nor sufficient to achieve higher-level efficiency or cost-effectiveness. Focusing on reducing costs can miss an opportunity to enhance VfM by spending slightly more and gaining disproportionately greater outcomes. Economy may be easier to measure than cost-effectiveness, and can be useful for programme management. However, care is needed to avoid a bias towards cost-cutting rather than value maximisation. Programme managers should maintain a line of sight on achieving efficiency at the highest possible levels of the results chain as this is a key factor in making judgements about whether a programme represents VfM.[9] Statements on VfM that rely on indicators at lower levels, as may be the case in early stages of programme implementation when outputs and outcomes have not yet been delivered, need to be duly qualified.

The terminologies used in FCDO's framework do not always align with common economic usage of terms. For example, 'efficiency' is applicable throughout the results chain, including inputs, outputs, and intermediate and final outcomes, whereas FCDO restricts this term to the output level alone (Renard and Lister, 2015). Even at output level, FCDO's definition of efficiency seems restricted to the concept of 'technical efficiency' (maximising outputs for a given level of inputs), whereas VfM evaluation could also usefully examine 'allocative efficiency' (which in an aid context is about allocating resources to the right mix of inputs to produce the intended outputs and outcomes) and 'dynamic efficiency' (e.g. learning, reflecting, adapting, responding to emergent opportunities, and becoming more efficient over time). In complex adaptive programmes, responsiveness to an evolving context may be more important for VfM than static concepts of how efficiently inputs are converted into outputs.

The term 'cost-effectiveness' may be confusing because it is also used to refer to a specific form of economic evaluation (cost-effectiveness analysis (CEA)), which measures the ratio between a single outcome indicator, measured in natural or physical units (e.g. lives saved),

[9] ICAI (2011: 1) reached a similar conclusion: 'In our view: effectiveness involves achieving a sustained impact for intended recipients, and value for money is the best use of resources to deliver the desired impact.'

and monetary programme costs. In VfM evaluation, a broader interpretation is needed: the 'cost-effectiveness' criterion focuses on the relationship between value created and value consumed, or whether sufficient outcomes or impacts are achieved to justify the investment of resources. Value created and consumed can potentially include both tangible (monetary, or readily monetisable) and intangible value (resources, outcomes or impacts that are difficult to value credibly in monetary terms).

4.3 Defining programme-specific criteria

Criteria need to be contextually determined. The generic definitions of the 5Es can be challenging to apply in practice because they are defined at a broad conceptual level that does not necessarily link in an intuitive way to the inputs, outputs, and outcomes of a specific programme or context. In order to support clear evaluative reasoning, we need to develop programme-specific definitions of economy, efficiency, effectiveness, cost-effectiveness, and equity that make these concepts directly relatable and meaningful to the programme at hand.

Furthermore, the 5Es do not define VfM. They are a useful checklist, but it might not be necessary or possible to apply all 5Es in some situations. Similarly, VfM could include additional criteria such as relevance (e.g. are we allocating resources to meet demonstrated needs?) or sustainability (e.g. how long does the value last?—see Section 4.7).

As can be seen from the example below, criteria are quite different from indicators in that they describe the relevant aspects of performance with reference to their intended functioning and effects, and not with regard to how they might be measured. While indicators are specific and measurable, criteria describe the nature of the changes that are intended. These descriptions are deliberately broader and less specific, reflecting their purpose of facilitating meaningful evaluative judgements.

Box 3: MUVA effectiveness criteria

As part of the MUVA female economic empowerment programme in Mozambique, effectiveness was broken down into three subcriteria, defined as follows (King and Guimaraes, 2016; King, 2019b).

Effectiveness as an urban female economic empowerment programme: MUVA's approaches make their intended contributions to capacity, agency, and/or opportunity as defined in project-level Theories of Change, and are scalable.

Effectiveness as a learning programme: Local participation, relationships, and knowledge contribute to project development; reflective learning processes are occurring as intended; and MUVA provides credible evidence about the effectiveness of every project, including evidence to enable decisions about which projects deliver better results.

Effectiveness as an influencing programme: MUVA is a recognised brand in the communities in which it is working; MUVA influences DFID's programmes in the wider Mozambique context; effective approaches are taken up and implemented by partners; and stakeholders become champions/agents of change.

4.4 Standards

Programme-specific criteria are important, but they are not enough to provide a transparent basis for distinguishing 'good' VfM from 'excellent' or 'poor' VfM. In addition to criteria, 'standards' need to be developed that specify 'what the evidence [would] look like at different levels of performance' (Davidson, 2014: 6). For example, ICAI (2011) provided a traffic light

rating system that represents a generic set of standards to support judgements about effectiveness and VfM.

We have developed the following set of standards to provide generic definitions of different levels of performance (Table 2). While the number of levels and description of the standards may be modified depending on the context of the intervention being assessed, we have found this framework generally works well in practice.

Table 2: Generic rubric to guide the calibration of performance standards

Performance standard	Definition
Excellent	The intervention is not only meeting all reasonable expectations/targets bearing in mind its context, but is substantively exceeding some of these. There may be room for incremental improvements.
Good	The intervention is generally meeting reasonable expectations/targets, allowing for a few minor exceptions. Some improvements may be needed.
Adequate	The intervention, though not meeting all expectations/targets, is fulfilling minimum 'bottom-line' requirements and is showing acceptable progress overall. Significant improvements may be needed.
Poor	The intervention is not fulfilling minimum 'bottom-line' requirements and/or not showing acceptable progress overall. Immediate and major improvements are needed.

Note: this colour palette is recommended to differentiate performance levels because it is designed to be accessible to individuals with colour blindness..

This generic rating system can potentially be used in conjunction with a list of programme-specific criteria to support evaluative judgements. However, it can be helpful to develop tailored standards to facilitate greater clarity in evaluative judgements. Programme-specific standards contain descriptions of relevant subcriteria describing what performance looks like at each level (as shown in Table 3). In so doing, such standards also indicate the relative importance of different criteria.

Although the following rubric is tailored to a specific programme, it is aligned with the generic definitions of 'excellent', 'good', 'adequate', and 'poor' performance from Table 2.

Box 4: Example of a programme-specific rubric for Economy

The following performance standards for Economy were developed for the independent evaluation of TradeMark East Africa (TMEA) Strategy 1 (Hansford *et al.*, 2019).

Table 3: TMEA Strategy 1 Economy standards

Performance standard	Definition
Excellent	TMEA can demonstrate that it has consistently maximised value in its procurement practices by following international best practice guidelines, drawing on multiple criteria which go beyond price alone, and ensuring that its partners follow the same practices. Also meets all criteria under 'good' performance.
Good	Average unit costs for key inputs generally meet agreed benchmarks. Indirect costs as percentage of total costs generally at or below agreed benchmark. Also meets all criteria under 'adequate' performance.
Adequate	Average unit costs for key inputs do not consistently or materially exceed agreed benchmarks. Indirect costs as a percentage of total costs generally near agreed benchmark. TMEA verifiably followed good practice to meet key economy drivers (staff remuneration, consultant procurement and fee setting, and procurement processes).
Poor	Any of the conditions for 'adequate' not met.

The following subsections provide practical guidance and tips for developing programme-specific criteria for interventions. Examples of programme-specific definitions and potential subcriteria are provided, illustrating how they are used in other VfM frameworks. This is intended to offer a menu of potential options, but it is not intended to be prescriptive or exhaustive, nor to constrain your use of the VfM approach.

4.5 Developing criteria and standards: three steps

Developing VfM criteria and standards involves the following three steps.

1. For each of the overarching criteria (economy, efficiency, effectiveness, equity, and cost-effectiveness), write a concise programme-specific definition
2. For each definition, identify subcriteria that describe the most important dimensions of good performance.
3. Organise the subcriteria into levels (e.g. definitions of excellent, good, adequate, and poor) in a table (or 'rubric').

4.5.1 A worked example: Efficiency, FSD Africa (FSD Africa, 2022)

Step 1: Write a concise programme-specific definition. This overarching definition is the highest-level description of the concept to be evaluated. When a performance rating (e.g. excellent, good, adequate, or poor) is made for efficiency, it is this overarching statement to which the rating will apply.

- *Efficiency: FSD Africa does the right things in the right ways, producing the intended quality and quantity of outputs within the available resources and timeframe.*

Step 2: For each definition, identify subcriteria that describe dimensions of good performance.

For FSD Africa, efficiency subcriteria were identified for (i) TA, partnerships, and networks; (ii) grants, returnable grants, and investment capital; and (iii) delivery of intended outputs, knowledge management, and data.

For brevity, this example shares details of the TA, partnerships, and networks subcriterion, which was further broken down into the following aspects, building on FCDO Guidance (FCDO, n.d.):

- TA, partnerships and networks are relevant, tailored to context, and complement other actors;
- TA is technically robust, up to date, and accessible; and
- TA supports partners to build their capacity at the organisation, process, or individual level.

Step 3: Organise the subcriteria into levels of performance. The levels are cumulative, so that the definition of 'good' includes all criteria for 'adequate', and so on (Table 4).

Table 4: Example of a programme-specific rubric for Efficiency

Performance standard	Definition
Excellent	TA is flexible and responsive to changes in context, with appropriate appetite to take calculated risks and work adaptively to gain traction. Also meets all criteria under 'good' performance.
Good	TA, partnerships, and networks are tailored to context and address needs that would not have been met by other actors (or would have been provided at lower quality). TA supports partners to build their capacity at organisation, process or individual level (e.g. knowledge, advice, mentoring; improvement of processes, practices or standards; support for strategy or policy development). Also meets all criteria under 'adequate' performance.
Adequate	Systems are in place to ensure TA, partnerships, and networks: a) are contextually relevant (i.e. not just transferring 'best practice' from elsewhere that will not work or is not realistic); and b) complement other actors. TA is accessible to relevant stakeholders at the right time. TA is technically robust and up to date (e.g. assessed through quality assurance processes).
Poor	Any of the conditions for 'adequate' not met.

The following should be considered when setting standards.

- It is often easiest to start by defining the 'adequate' and 'poor' levels (i.e. the difference between a pass and a fail in VfM terms), then the 'excellent' level (the highest aspirations), before interpolating standards for the 'good' level.

- While all levels need to have unambiguous definitions to support clear judgements, it is especially important to have explicit 'fail criteria', delineating the boundary between what is 'good enough' and 'not good enough' VfM.
- Some subcriteria might belong at one level only, e.g. they might form part of the definition of 'adequate' or 'excellent'. Other subcriteria might have a range of definitions at different levels, e.g. the same dimension of performance might be described in a different way at each level.

4.6 Potential content of criteria

The following subsections provide further detail on the potential content of VfM criteria, with a view to promoting the quality and consistency of VfM frameworks. It is important to stress that criteria need to be contextually defined. The following considerations are not exhaustive, and are not intended to prescribe or constrain programme-specific VfM criteria.

4.6.1 Economy

Economy is concerned with the conversion of resources (e.g. money) into the inputs (products, services, staff, assets, etc.) required to deliver an intervention. Historically, definitions of economy have focused on frugal management of donor funding, spending less and minimising the cost of inputs. Reducing the costs of an intervention can improve an intervention's VfM *if* the intervention can maintain the effectiveness and equity of the value it delivers. However, the challenge with defining economy based on cost minimisation only is that it can lead to indicators and incentives that are counterproductive for an intervention's value proposition. Put another way, the cheapest option is not always the best option.

Most definitions of economy reflect this principle and highlight the importance of the types and quality of inputs purchased by an intervention. For example:

- NAO defines economy as 'Minimising the cost of resources used while having regard to quality' (NAO, n.d.); and
- FCDO defines economy as 'Are we (or our agents) buying inputs of the appropriate quality at the right price?' (DFID, 2020).

FCDO's definition of economy is not about buying what is cheapest, but about using resources as effectively as possible—including financial, political, and human resources, as well as costs to recipients.

Therefore, a programme-specific definition of economy could take the following form:

> *The XYZ Programme team is a good steward of programme resources, buying inputs of the appropriate quality at the right price.*

Potential subcriteria of economy could include considerations such as:

- paying reasonable prices for inputs (e.g. average costs of significant items such as daily TA consultancy fees, air fares, and per diems);
- using good procurement practices (e.g. competitive tendering where appropriate and feasible, whole-of-life costing of significant items, planning in advance where possible);
- managing risks of cost increases (price- or volume-related);

- proactively finding economies of scale/scope, savings and 'best deals' to manage costs down;
- leveraging support from partner organisations to 'grow the funding pie';
- sound financial management and probity arrangements; and
- risk management to minimise frictional losses, e.g. potential leakage due to informal or corrupt transactions (after funding has been disbursed from the programme to subcontractors or the community).

In summary, economy typically involves looking at an intervention's main cost drivers and cost control processes, and will involve assessing financial indicators. Considering quality and appropriateness of prices within the relevant economy can help steer us towards more meaningful economy indicators, such as intelligent procurement practice or robust fiduciary risk management.

Box 5: Defining economy: Going beyond cost minimisation

Economy often focuses on purchasing and procurement and is expressed in financial terms about cost control and cost minimisation. However, good stewardship of resources goes beyond management of programme funding. For example, donors and programme leaders increasingly recognise that programmes should account for their impacts on natural systems.

For instance, the purchase of inputs for interventions can lead to negative externalities such as increased carbon emissions. Interventions might be able to reduce carbon emissions through minimising air travel or developing procurement processes that give preference to local suppliers, etc. Conversely, there may be potential to support positive externalities, such as societal benefits that may stem from a programme using procurement to help build the capacity of local suppliers.

Personnel is often a key cost driver for interventions, and staff are often among the most valuable and important inputs of an effective programme. Good stewardship of resources includes considering issues such as equitable pay and sourcing staff from diverse backgrounds, which can make a broader contribution to programme effectiveness and wider social issues.

These issues may be overlooked in VfM analysis because they may, although not necessarily, lead to higher costs of inputs, and are not reflected in the value propositions expressed in business cases. However, they often warrant consideration as economy targets with an over-reliance on cost minimisation can lead to reduced value creation overall.

4.6.2 Efficiency

Efficiency, as defined by FCDO, addresses the question:

How well are we (or our agents) converting inputs into outputs?

FCDO's definition of efficiency is aligned with the concept of *technical efficiency* (maximising output for a given level of input). However, it is often helpful to consider other aspects of efficiency as well.

The 'technical efficiency' perspective is often important when assessing the performance of implementers, when the VfM evaluation has an accountability focus. For example, imagine a project has an output target to deliver training to 1,000 health workers for US$50,000 and contributes to an outcome of improved patient care. A VfM evaluation finds the project has delivered training to 1,100 health workers for US$50,000, but there is no evidence that the project has contributed to improved patient care. Perhaps the supplier focused on the target at the expense of the quality of the training, or perhaps the supplier did a good job but the training design was the wrong one. This is a good illustration of why we should not rely on

technical efficiency (or indicators) alone. To understand the efficiency of the project, we need to take additional factors into account.

For the VfM evaluation of larger programmes or policies, there is generally an expectation that the initiative will lead to improvements in outcomes as well as outputs. The expectation on implementers extends from expecting them to deliver outputs on time, to cost, and to agreed quality standards, so that they will manage resources effectively to achieve the intended outcomes. In these cases, an input-to-output definition of efficiency can become inadequate. Additional aspects of efficiency, such as allocative efficiency, dynamic efficiency, and relational efficiency, can be helpful as they enable a deeper assessment of whether programmes are delivering productively.

Potential subcriteria for output efficiency or productive delivery include the following.

- **Technical efficiency**, or maximising productivity by using the least resources necessary to deliver the required quality and quantity of outputs: We can flip this around and see how well a programme is delivering the overall volume of work expected within budget, on time, and to the agreed quality standard. This is often more complex and nuanced than simply meeting planned outputs. In complex, adaptive programmes, we would expect to see plans change—so we need to guard against penalising programmes if delivery does not match original plans. We can use Mintzberg's 'Emergent Strategy' framework as a way of tracking how a programme evolves (Figure 3).
- **Allocative efficiency**, or the right mix of inputs and activities to maximise productivity: In economic theory, allocative efficiency occurs where the distribution of goods and services in the economy matches consumer preferences and price is equal to the marginal cost of production. In public services, we can use the concept analogously to refer to allocating resources to the right mix of interventions to meet needs or objectives.
- **Dynamic efficiency**, or adapting and improving: In economics, dynamic efficiency refers to the idea of optimising resource allocation over time so that no generation can be made better off without another being made worse off. In more practical terms, we can look at how well a programme adapted and improved its productivity over time, through evaluating, reflecting, learning, adapting, adopting new technologies, responding to emergent opportunities, etc. Trends in technical and allocative efficiency over time can also help assess dynamic efficiency.
- **Relational efficiency**: Relationships, communication, and trust are a foundation that enables programmes to operate efficiently. Without them, resources may be wasted. A new programme may become more efficient (and effective) over time as relationships develop and it becomes a trusted part of the landscape. Building relationships requires investment and should be made explicit in order to understand VfM.

An example of a programme-specific definition of efficiency, reflecting these considerations, was TMEA. In its VfM evaluation, the overarching definition of efficiency was:

> *The TMEA programme produced the intended quality and quantity of outputs within the available resources and optimised the use of resources by moving resources around for greater leverage.*

This programme-specific definition of efficiency was further detailed in subcriteria, including following good practices to manage key drivers of technical efficiency; delivery of programme outputs (taking adaptive management into account); use of economic analysis and systematic project appraisal processes to inform sound investment decisions (allocative efficiency); responsiveness to context, opportunities, and challenges/risks; and ongoing learning to leverage higher returns (dynamic efficiency) (Hansford *et al.*, 2019).

Emergent strategy

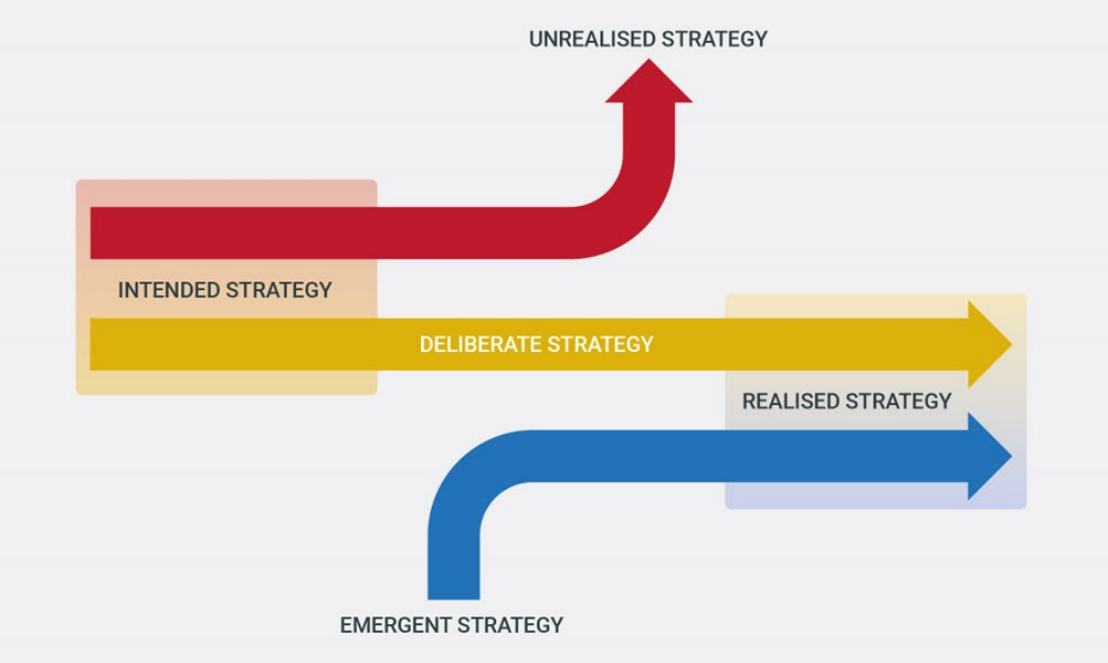
Complex development programmes, working in complex political and market contexts, do not operate in a linear fashion. This influences the way we need to look at VfM. It would be inappropriate to draw simple links between activities and outcomes, without due consideration of the complexity of the programme and the context in which it operates.

There is a body of literature on how complexity approaches differ from more linear management approaches (Olson and Eoyang, 2001).

- There may be multiple actors, and the behaviour of the system can be unpredictable and uncontrollable (rather than stable, predictable, and controllable).
- Direction-setting, when attempting to bring about system change, requires participatory approaches, often with multiple stakeholders (rather than taking a top-down, directive approach).
- Causality tends to be non-linear, with multiple variables and interrelationships (rather than being linear with every effect traceable to a specific cause).
- Responsiveness to the environment is an important measure of value (with simple concepts of input-output efficiency and reliability being less relevant) (Judy Oakden, 2017, personal communication).

Accordingly, development programmes must be responsive to their evolving context. There will be some aspects of strategy that are planned ('intended strategy') but that do not take place due to the evolving context ('unrealised strategy'). At the same time, new approaches ('emergent strategy') will be adopted. Valid judgements about VfM will need to recognise this feature of complex development programmes, determine whether adaptive management is occurring and is effective, and account for unrealised and emergent strategy. This is illustrated in Figure 3.

Figure 3: Emergent strategy



Source: Based on Mintzberg and Waters (1985)

The box below briefly describes how the VfM evaluation of the Pakistan SNG Programme was able to take account of, and reflect, emerging strategies (King and Allan, 2019).

Box 6: The VfM evaluation of SNG: A complex governance programme

SNG was a complex governance reform programme. Its complexity was derived in part from its operating environment, which necessitated a keen understanding of the prevailing political economy in Punjab and Khyber Pakhtunkhwa, including the dynamic power structures within the two provincial governments and their changeable relationships with the central and district governments during a period of considerable decentralisation reform. The VfM evaluation needed to be cognisant of this context, and needed to assess the programme's relevance in a complex and rapidly changing environment.

An additional challenge in assessing the VfM of SNG was that the value of the programme came not only from the achievement of the specified outcomes, but also critically from its ability to be adaptive, responding to lessons learnt as well as emergent opportunities and challenges. Accordingly, the VfM framework was designed so that it would not unduly penalise the programme if some of its interventions were found to be ineffective and were subsequently disbanded, as this was inherent to the programme's adaptive design. Finally, the flexible and responsive nature of SNG meant that many of the programme's outputs could not be defined at the outset. As such, the benchmark against which effectiveness and efficiency would be assessed also needed to be sufficiently flexible. For example, revenue support was not in the original SNG logframe, but it was quickly adopted by the programme when the Government of Khyber Pakhtunkhwa expressed an interest in supporting its new revenue authority during the 2013 Inception Phase.

4.6.3 Effectiveness

Effectiveness is about outcomes – an important dimension of VfM. FCDO's approach to VfM asks:

> *How well are the outputs produced by an intervention having the intended effect?* (DFID, 2020)

A programme-specific definition of effectiveness could take the following form:

> *The XYZ programme achieves its intended outcomes within the available resources.*

If possible, a programme's outcomes should be expressed in more specific terms, aligned with its ToC. For example, in the Pakistan SNG Programme, the effectiveness definition was:

> *The SNG Programme achieves its intended changes in public financial management, governance and planning systems, and service improvement pilots, in Punjab and Khyber Pakhtunkhwa.* (King and Allan, 2018)

Subcriteria of effectiveness will refer specifically to outcomes identified in the ToC. Therefore, examples of potential subcriteria are not included here. These will often include intermediate outcomes that indicate the direction of travel toward longer-term outcomes, e.g. changes in knowledge or behaviour rather than changes in health status or educational attainment. Note that logframe 'outputs' are sometimes better classified as lower-level outcomes for evaluation and VfM purposes. For clarity, outputs are products delivered by the programme and substantively within its control, whereas outcomes involve some action or behaviour on the part of people who are not directly involved in delivering the programme and who are within the influence, but not under the direct control, of the programme.

Often, intended outcomes are identified by programme architects. However, we may also consider recipients' needs and expectations of the programme. Some effects may be unintended. Unintended outcomes may be positive or negative and should be sought for while assessing effectiveness. Considering questions like effective 'for whom' and 'in what circumstances' also matters as communities and individual people may experience unique outcomes. It is therefore important to understand varying outcomes.

4.6.4 Equity

FCDO defines 'equity' as spending fairly, with due consideration to the different constraints and opportunities people face and how these may affect their ability to achieve different outcomes (DFID, 2019).

> *How fairly are the benefits distributed? To what extent will we reach marginalised groups?* (DFID, 2020)

At a deeper level, the fundamental purposes of many aid and development programmes include addressing inequities—between countries or regions, and/or between subgroups of people in those countries/regions. Some of these programmes only exist because inequities exist. In these contexts, equity needs to be given due prominence in VfM evaluation. In some cases, it may be deemed appropriate to position equity as the 'first E', paving the way for a framework to consider what it means to 'deliver equity gains efficiently and effectively'. This opens the possibility of using VfM evaluations to add to the body of knowledge about what it costs to deliver programmes that are effective at addressing inequities.

Equity can be assessed at programme design through to implementation and closure. It is therefore important that equity should be considered alongside economy, efficiency, and effectiveness. Equity considerations are thus embedded throughout the VfM framework. Nonetheless, it is important to include an explicit criterion for equity so that judgements of VfM give sufficient prominence to the targeting of resources and results, e.g. to poor people, women and girls, disabled people, and other marginalised groups.

An example of a programme-specific definition of equity is:

> *The XYZ programme reaches and benefits its intended target groups.*

Where possible, this should be made more specific with reference to the intent of the programme, aligned with the ToC. For example, in the SNG Programme, equity was defined as follows:

> *Changes in needs-based planning and resource allocation contribute to reducing inequities by targeting resources to poor people, women and girls.* (King and Allan, 2018)

Potential subcriteria of equity could include:

- equitable design, e.g. programme design and needs assessments explicitly identify target groups and appropriate strategies for reaching them;
- equitable access, e.g. interventions are accessible and acceptable to people from key target groups (and are accessed by those groups), and frictional losses (such as interventions benefiting people outside the target groups) are minimised;

- equitable delivery, e.g. project appraisal and investment decisions explicitly allocate resources to key target groups, implementing explicit strategies to reach target groups, and monitoring results to understand who benefits from the programme, including the views of recipients and their representatives; and
- equitable outcomes, e.g. outcomes improve for the key target groups.

4.6.5 Cost-effectiveness

According to FCDO, cost-effectiveness considers the relationship between impact and total costs incurred:

> *What is the intervention's ultimate impact on poverty reduction, relative to the inputs that we or our agents invest in it?* (DFID, 2020)

The key principle underlying the cost-effectiveness criterion is that the outcomes achieved by a programme should be 'good enough' (or better than good enough), bearing in mind the level of investment in the programme. FCDO's definition is therefore *not* directing us to do a CEA, or any other form of economic analysis such as a cost–benefit analysis (CBA). It is prompting us to think about the relationship between value created and value consumed. Economic methods of evaluation are often an appropriate way to inform a judgement on the criterion of cost-effectiveness but they are not the only way. Sometimes it may not be feasible or desirable to use economic methods (see Annex A). **Ultimately, what matters is that we determine whether enough value is created to justify the resource use**.

Therefore, an example of a programme-specific definition of cost-effectiveness could be:

> *The outcomes/impacts of the XYZ programme are commensurate with the level of investment.*

This implies that a programme should create more value than it consumes, or should provide a positive return on investment (ROI). However, in aid programmes, the reality may involve accepting that a programme creates less value overall than it consumes (in purely economic terms) in improving equity for marginalised groups of people. Accordingly, the cost-effectiveness criterion requires a well-reasoned judgement and cannot necessarily rely solely on economic analysis. Unpacking a programme's value proposition by considering how the programme creates value, what kinds of value it creates, and for whom it is created (see Section 3.1), can help in identifying criteria to determine whether a programme creates enough value to justify the resources used.

Potential strategies for addressing cost-effectiveness, which could form the basis for developing programme-specific sub-criteria, include:

- impacts/outcomes and their associated costs are comparable to those seen in other relevant interventions in the same or similar contexts (this could be informed by specific indicators such as cost per quality-adjusted life year or a more qualitative assessment of other programmes);
- impacts/outcomes and their associated costs are justifiable relative to the costs of 'doing nothing' (e.g. burden of disease);
- the value of impacts/outcomes achieved to date exceeds the value of resources invested (e.g. CBA);
- the estimated value of future impacts/outcomes (projected on the basis of early results and judgements of their sustainability, over a reasonable time horizon and at a suitable

discount rate) is likely to exceed the value of resources invested (e.g. a break-even analysis);
- the value of impacts/outcomes achieved to date, relative to the value of resources invested, is within the acceptable or anticipated range given the impacts achieved on equity;
- programme expectations are met for the level of funding allocated (i.e. in a fixed-budget context, cost-effectiveness can justifiably be deemed equivalent to achieving 'enough' outcome/impact);
- where a programme ToC sets out short- and longer-term outcomes (the latter may be referenced as higher-level outcomes, impact, systemic change, transformational change, etc.), it may make sense to assess progress towards short-term outcomes as part of Effectiveness, and longer-term outcomes as part of Cost-effectiveness.

4.7 Additional VfM criteria

If we define VfM as *good resource use*, we need to consider whether there are additional aspects of good resource use beyond the 5Es in a given programme and context. 'Good resource use' is a multidimensional concept, so in some contexts it may make sense to use other criteria to supplement, or replace some of, the 5Es. Examples of other possible criteria include sustainability, relevance, and coherence (in line with the Organisation for Economic Co-Operation and Development (OECD) Development Assistance Committee (DAC) Criteria for Evaluating Development Assistance),[10] scalability, and acceptability/satisfaction. The OECD recommends that criteria should 'provide complementary perspectives, giving a holistic picture of the intervention' and 'the use of evaluation criteria should be thoughtful and adapted to the purpose' (OECD, 2021).

This is not intended to imply that all of these criteria should be used in every VfM evaluation. The appropriate selection of criteria will be determined by the particular needs and context of a programme, stakeholder interests, and whether the VfM evaluation is situated within a broader evaluation or MEL system with its own criteria (Section 5). The resources available for VfM evaluations, and the need to keep VfM evaluation commensurate with the scope of the programme, are also important considerations in determining the number and scope of criteria included in a VfM framework. The '80/20 rule' described in Section 3.2 can be useful for keeping the criteria manageable.

The following principles may guide decisions on the selection of appropriate criteria (King, 2023b):

- **pragmatism:** the organisation commissioning the VfM evaluation may already have a view of what they mean by VfM;
- **manageability:** sticking with a manageable number of criteria that together capture the most important aspects of the programme (the 80/20 rule);
- **complementarity:** MEL and VfM frameworks should minimise duplication of content—for example, in areas of overlap (such as effectiveness) the VfM report could cross-reference the outcome evaluation;
- **conceptual coherence:** MEL and VfM frameworks should be consistent and non-contradictory (e.g. they should use a shared ToC); and

[10] 4. Understanding the six criteria: Definitions, elements for analysis and key challenges | Applying Evaluation Criteria Thoughtfully | OECD iLibrary (oecd-ilibrary.org).

- **practical alignment:** MEL and VfM activity should be coordinated to minimise duplication of activity (e.g. conduct a shared survey).

Note that rubrics are not just a VfM tool. One way to promote coherence between VfM evaluation and wider MEL activity is to use rubrics in a consistent way across the entire evaluation to support evaluative judgements about merit, worth, or significance (i.e. the 'how good' questions).

4.7.1 Sustainability

Sustainability is one of the six OECD DAC evaluation criteria (OECD, 2021) and is sometimes included as a VfM criterion. Sustainability is defined as 'the extent to which the net benefits of the intervention continue, or are likely to continue'. If sustainability is integral to the value proposition, then assessing sustainability can help us understand how much value a project, programme, or policy has delivered, or is likely to deliver in the future. This can in turn help us determine whether an initiative is cost-effective. For example, if a CBA shows that a community project will only deliver sufficient value after operating for 10 years of net benefits, then understanding if the project has the necessary institutional and financial capacity to be sustained for 10 years is important for assessing its overall VfM.

A key point when deciding whether sustainability is a worthwhile criterion to include is whether it is integral to the value proposition. Programmes that are sustainable are not necessarily better VfM than those that deliver short-term benefits.

4.7.2 Relevance

Relevance is another OECD DAC criterion that can be applicable in VfM evaluations. OECD DAC defines relevance as 'the extent to which the intervention objectives and design respond to recipients, global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change'. In VfM terms, we can ask: are resources being allocated to relevant needs and priorities?

Decisions about relevance are made when considering the business case for a new intervention. However, population needs can change, and evaluations can uncover new knowledge that throws new light on needs and priorities. A key point when determining whether relevance should be included as a VfM criterion is whether there have been material shifts in context that might warrant a review of resource allocations to the programme.

4.7.3 Coherence

OECD DAC defines coherence as 'the compatibility of the intervention with other interventions in a country, sector or institution'. The coherence of a policy or programme is an important determinant of whether it represents good use of resources. It may therefore sometimes be important for VfM evaluations to consider internal coherence ('synergies and interlinkages between the intervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres'), and/or external coherence ('the consistency of the intervention with other actors' interventions in the same context', e.g. adding value while avoiding duplication of effort).

A key point when considering whether coherence may be an important criterion of VfM is whether stakeholders have questioned the coherence of the programme, or whether anything has changed in the operating environment that might affect its coherence (such as a change of government, or the establishment of a significant new programme with overlapping objectives). If so, it may be worth reviewing coherence to check whether the programme still represents good use of resources.

4.7.4 Scalability

Some programmes involve piloting interventions to determine whether they are effective, and then scaling them up. In these cases, the VfM of the programme is related (at least in part) to its success in piloting effective, scalable interventions.

Scalability depends on multiple factors, which may vary depending on context. For example, in the MUVA programme, we assessed scalability based on five considerations:

- **evidence of outcomes** (how effectively pilots achieved positive outcomes);
- **beneficiary buy-in** (whether the interventions responded to real problems that people experience and whether there was demand from recipients for the solution);
- **cost at scale** (whether the intervention could be affordable and cost-effective at scale);
- **relevance** (whether there was alignment between the objectives of donors and governments, and whether there was appetite and funding to drive the approach forward); and
- **reach** (the potential to reach large numbers of recipients at scale).

When assessing the potential cost-effectiveness of a pilot at scale, it is important to note that the ROI of a pilot is a poor and misleading predictor of ROI at scale. However, pilots can provide important clues about the nature and structure of costs and benefits, which can be used to inform analysis of potential cost-effectiveness at scale.[11]

4.7.5 Acceptability/satisfaction

Independently of whether a programme delivers outputs and achieves outcomes, is the programme acceptable to recipients and other stakeholders? In some instances, understanding acceptability and satisfaction is essential to understanding why a programme works or does not work, whether it is suitable for scaling, or what changes are needed to make it suitable. Therefore, understanding acceptability and satisfaction can be crucial to understanding whether a programme is a good use of resources (i.e. VfM).

[11] For details, see www.julianking.co.nz/vfi/pilot-roi/.

5 Integrating VfM with evaluations and MEL systems

Throughout this guide, we have mainly discussed VfM design and implementation in the context of standalone VfM evaluations. However, VfM can be one component of an independent evaluation, or it may be an integral part of a broader MEL system. When this is the case, careful thought is required to integrate VfM as far as possible and to avoid creating separate systems that result in duplication of effort (e.g. in data collection; reporting). We address the challenges of integrating VfM with, first, evaluations, and second, MEL systems in the following subsections.

5.1 Aligning VfM with other evaluations

Increasingly, VfM is included as a component of independent evaluations, often as an additional evaluation criterion or as a key evaluation question. VfM can be a valuable addition to an evaluation, adding questions that are often ignored in evaluations, such as how well resources are (or were) used; what value has been created; and whether the value created justified the financial costs and the opportunity cost. These questions can add value in different ways to different kinds of evaluations—formative and summative; and process, outcome, and impact evaluations.

We have faced a number of challenges in our work to integrate VfM with broader evaluations. These are described below, along with some strategies for dealing with the challenges.

5.1.1 Conceptual challenges (design)

VfM has often been treated as if it is separate from other elements of evaluation, with different reports written by different people, sometimes for different audiences and purposes. However, when we define VfM as *good resource use,* we acknowledge that VfM assessment *is* evaluation. It involves answering evaluative questions about resource use, such as: how well have resources been used? What value has been created? Does the value created justify the cost? How can more value be created from the available resources?

Conceptually, we can regard VfM as something that intersects with evaluation, something that serves evaluation, and something that is served by evaluation. Moreover, VfM criteria and other evaluation criteria overlap in various ways. For example, one set of evaluation criteria, the OECD DAC criteria, include relevance, coherence, effectiveness, efficiency, impact, and sustainability. Two of these (effectiveness and efficiency) overlap with the 5Es. However, the situation is more complex, because the OECD DAC definition of efficiency spans three of the 5Es (economy, efficiency, and cost-effectiveness), and because relevance, coherence, and sustainability may also be pertinent to good resource use (King, 2023b). Pragmatically, boundaries must be drawn, and this will often be determined by client needs. For example, VfM evaluations for FCDO generally address the 5Es.

Ideally, an evaluation will use a single framework that incorporates VfM-specific criteria alongside criteria for the broader evaluation. This requires understanding the criteria set out within an evaluation framework and how they relate to the VfM question posed, and then discerning whether additional VfM criteria are needed. Sometimes this may be quite

straightforward. For example, where an evaluation looks at performance across the ToC in such a way as to help understand what value is being created (delivery, outputs, outcomes, and impact), it may be sufficient to add VfM criteria to assess how well resources are being used (subcriteria to address 'Economy'), and whether the value created was worth the financial and opportunity costs (subcriteria that address 'Cost-effectiveness'). In other cases, there may be a need to align all of the criteria to arrive at a coherent framework. It all depends on the programme and context, the interests of key stakeholders, and the focus and scope of the evaluation. Table 5 demonstrates how evaluation criteria (both the 5Es and OECD criteria) can be mapped to core VfM questions.

Table 5: Aligning evaluative criteria

VfM questions	Criteria (5Es, OECD/DAC, other)
1. How well have resources been used?	Economy
	Efficiency (technical, allocative, dynamic, relational)
	Relevance
	Coherence
	Risk
	Affordability
	Ethics
	Equity (inputs)
2. What value has been created?	Effectiveness (outcomes)
	Equity (outputs and outcomes)
	Impact
	Scalability
3. Does the value created justify the cost?	Cost-effectiveness
	Sustainability
	Equity (impacts)

5.1.2 Methodological challenges (design)

Our approach to VfM evaluation recommends using evaluative reasoning based on pre-agreed rubrics to make evaluative judgements. Evaluative reasoning involves using an explicit and transparent approach to derive evaluative conclusions from evidence and show the reasoning in our reports. Evaluative reasoning is key to good evaluation practice (Yarbrough *et al*., 2011), although rubrics are only one way to do it. Alternatives to the use of rubrics include multicriteria decision analysis, if-then statements, deliberative, all-things-considered, and tacit approaches (Schwandt, 2015). These options are not mutually exclusive. We can combine them, e.g. using rubrics to guide deliberative approaches with stakeholders, or using tacit judgements as a validity check (King, 2023c).

We encourage the use of rubrics to make evaluative judgements (addressing the 'how good' questions) across the whole evaluation. However, if the broader evaluation does not include the use of rubrics, we recommend at least using rubrics to address the criteria that are relevant to VfM. This is a core feature of the OPM approach to VfM evaluation.

Some evaluations focus on assessing or measuring causal relationships, particularly the specific question of whether an intervention has had an impact on defined outcome indicators and how large that impact has been. In this case, the use of rubrics and the VfM component of the evaluation can help make a judgement on the value of that impact, taking into account the perspectives of different stakeholders, a question that may sometimes be missed in some narrow impact evaluations.

5.1.3 Practical challenges (implementation)

Where VfM and evaluation are not well-aligned, this can result in duplication of effort. For example, the two frameworks might include similar data to be collected according to slightly different timeframes. Or they might require different framing of the same data, leading to two interpretations and presentations of the same data through different lenses. Ultimately, if the two designs do not cohere conceptually and methodologically, we may end up with parallel exercises of data collection, data analysis, and reporting that use some of the same, or similar, data. This can be wasteful in terms of the effort dedicated to the evaluation, and demanding on stakeholders who are forced to make sense of the same, or similar, information presented in divergent formats.

To avoid these problems, develop the evaluation framework, VfM framework, and associated methodologies together during the inception/design phase. This can be facilitated by following the eight-step process set out in this guide, summarised as follows.

- **Step 1: Understand the programme and define a coherent approach to the evaluation:** This includes identifying how the different objectives, questions, and work streams (e.g. process, outcome, VfM) will complement one another, agreeing boundaries between the VfM component and other components of the evaluation, negotiating a coordinated reporting approach, and developing a shared ToC.
- **Step 2: Develop criteria:** This includes defining evaluation criteria and VfM criteria, the boundaries between them, and how to manage overlaps. For example, to avoid duplication, the Effectiveness section of the VfM report could cross-reference the outcome evaluation.
- **Step 3: Develop standards:** Define what 'good' looks like for *all* criteria (not just VfM), as suggested under the methodological challenges above.
- **Step 4: Identify evidence sources and methods:** This includes identifying what data the evaluation will provide and what additional data are needed for the VfM component (e.g. cost data; data from an economic evaluation, if appropriate). It also includes identifying opportunities to serve evaluation *and* VfM purposes jointly and in a coordinated way, e.g. shared surveys, key informant interviews, etc.
- **Step 5: Gather evidence:** This includes coordinated management of evaluation and VfM activity to avoid unnecessary duplication of effort.
- **Step 6: Analyse evidence:** During this stage, different team members may be analysing different streams of evidence separately. However, those responsible for VfM evaluation will need to liaise with other evaluation team members, as needed, to ensure the different streams of evidence contribute appropriately to the VfM evaluation.
- **Step 7: Synthesis and judgements:** This involves using the rubrics to interpret the evidence and judge performance, quality, and VfM on an agreed basis. Ideally, this will use a consistent approach for all criteria, as suggested under the methodological challenges above.
- **Step 8: Reporting:** This includes preparing and delivering a coordinated, coherent set of evaluation and VfM reports, or, better, a single report that integrates findings from the

evaluation and VfM evaluation into a single coherent story. Where the VfM report cross-references other MEL deliverables, as is frequently the case (e.g. process evaluation, outcome evaluation, gender equality and social inclusion (GESI) evaluation, or Annual Performance Report), staggered timeframes should be negotiated to allow the VfM team to work from the final versions of the relevant MEL reports. Sufficient time must be built in to extract the necessary evidence, reframe it, and draw conclusions about VfM.

It is easier to fully integrate VfM with the evaluation if at least some (if not all) team members work on both the overall evaluation and its VfM component. If this is not feasible, there should be someone to oversee and coordinate the full evaluation and ensure that the different components remain aligned.

5.2 Aligning VfM frameworks with programme MEL and routine reporting

Integrating VfM evaluation with programme MEL systems is good practice, since programmes should be regularly judging whether they are delivering good value and looking for opportunities to improve it as part of a wider process of reflection and learning. Some donors (such as FCDO) will expect programmes to report against a VfM framework, and incorporating this into the MEL system will help ensure coherence between them and avoid duplication and inefficiencies. Ideally this will be done as part of designing the MEL system. The organisation and coding of financial data (e.g. by workstream) should also be considered at this point.

As noted earlier (Section 2.2), pragmatic decisions are required to define the scope of a VfM evaluation to complement wider MEL work, to align them conceptually (e.g. based on the same ToC), and to coordinate the work programme.

The approach, as laid out in Section 3, recommends that the identification of evidence requirements should only be carried out after the criteria and standards are developed to ensure that evidence is relevant to the aspects of performance being assessed. In practice, this can be challenging when developing VfM frameworks for large programmes or organisations where there are pre-existing reporting mechanisms against MEL frameworks and/or contractual KPIs. Adding requirements for VfM data collection and reporting, where this has not been integrated from the beginning, can be costly and perceived as burdensome. Where this is the case, it can be helpful under 'Step 1: Understand the programme' to assess existing evidence sources and reporting requirements. This helps us leverage existing sources of data and reduce the time and resources needed to perform VfM evaluations in future. Table 6 summarises where programmes may have existing reporting requirements that could be used to develop VfM criteria, standards, and evidence requirements.

Table 6: Evidence sources when developing programme VfM frameworks

VfM criteria	Possible evidence sources	Common areas of alignment
Economy	**Contract/ financial KPIs**	Programmes often have contractual or financial KPIs that ensure suppliers manage and minimise costs. Programme managers or finance managers can often provide details of financial KPIs that may also be relevant for assessing economy performance. For example, financial KPIs might include: • average daily fee rates do not exceed US$xx; • audits demonstrate strong financial compliance; • budget variance (actual vs forecast) is +/- 5%; • annual budget utilisation >95%; or • at least 25% co-funded from other sources. Some financial KPIs can contribute towards understanding whether a programme is converting resources into inputs effectively (i.e. good economy). For example, measuring average fee rates can help us assess whether a programme is paying competitive rates for staff, or successful audits can demonstrate that programmes have strong financial procedures that reduce fiduciary risks. However, financial KPIs are not always appropriate as economy indicators, and just because they are readily available does not mean they should be used. For example, accurate budget variance or high levels of budget utilisation may be important for funders' own financial management objectives, but these seldom drive VfM performance on programmes and can be detrimental to good overall VfM. For example, restricting budget variance can limit the flexibility of a programme to make changes, and budget utilisation targets can incentivise programmes to focus on spending their available budgets at the expense of a slower and more cost-effective approach. Likewise, co-financing is often an important KPI for funders to ensure that other partners are making relevant financial or non-financial contributions to a programme. While it may appear intuitive that additional resources contribute to good economy, this cannot be taken for granted. For example, if a Ministry of Health provides US$1 million of funding for the delivery of a health programme, there is an opportunity cost to this (i.e. the Ministry of Health is spending US$1 million less on something else that delivers value). The critical question to ask when assessing available indicators is whether the resources are well-used to deliver more value overall.
Efficiency	**Progress reports/ logical frameworks**	Where efficiency is defined as how well inputs are being converted into outputs, progress reports are often a key source of evidence. Particularly where VfM evaluations are accountability-focused, progress reports can be useful for establishing whether deliverables have been achieved or whether other performance targets have been met.

VfM criteria	Possible evidence sources	Common areas of alignment
		Other forms of efficiency, such as dynamic efficiency or allocative efficiency, tend to require more investigation. However, progress reports often still provide details on resource allocations, lessons, and adaptations.
Effectiveness	**MEL frameworks/ logical frameworks**	Effectiveness is principally concerned with the achievement of outcomes. As described in Section 4.6.3, it can be helpful when developing the effectiveness criterion to break it into subcriteria to reflect different intended outcomes. Most programmes have existing documentation that lay out the intended outcomes (such as a logical framework, results framework, ToC, or MEL framework), and it is normally appropriate for the effectiveness criterion to be aligned with these outcomes. For example, if a programme has three intended outcomes in a logical framework, it is often logical for the VfM effectiveness criterion to have three subcriteria that mirror these. Sometimes MEL frameworks can focus only on measurable outcomes, at the expense of value, which is where a value proposition may be useful. However, MEL frameworks are a key starting point when looking at effectiveness criteria.
Equity	**MEL and GESI frameworks**	Some programmes have specific approaches or frameworks for assessing GESI, which can often be utilised. Equity is very context-dependent, and existing programme definitions of the types of group that are being targeted can often help develop criteria and standards. In the absence of a GESI framework or similar, programme MEL data are often disaggregated by gender and other characteristics.
Cost-effectiveness	**Economic appraisals**	Developing programme-specific criteria and standards for cost-effectiveness is often one of the most challenging criteria to complete. Often programme finance teams do not have financial data structured by output or outcome, which restricts cost analysis, and MEL teams often do not have the data necessary to use economic methods of evaluation. A useful starting place is often the programme business case or design documents. Most funders require some form of economic appraisal during the design and development stages of programmes, and these can often provide a basis for assessing cost-effectiveness during implementation. For example, FCDO programmes must complete a business case, including an economic appraisal that identifies and values the relevant costs and benefits of the proposed programme before they are approved. Similarly, the Millennium Challenge Corporation and Asian Development Bank require a CBA analysis before approving programmes to demonstrate that the minimum economic rate of return is 10%[12] and 9% respectively.[13]

[12] Economic Rates of Return (mcc.gov).
[13] Guidelines for the Economic Analysis of Projects (adb.org).

VfM criteria	Possible evidence sources	Common areas of alignment
		Economic appraisals carried out during the design stages of programmes can often guide the definition of programme-specific cost-effectiveness criteria, provide benchmarks for what good performance may be, and determine what methods of economic evaluation may be appropriate. However, economic appraisals are not always the most appropriate way to judge cost-effectiveness, and due consideration should be given to the programme's overall value proposition and the context (Section 4.6.5).

Bibliography

Adler, M.D. and Posner, E.A. (2006) *New Foundations of Cost–Benefit Analysis*, Harvard University Press, Cambridge, MA.

Bamberger, M. (2012) 'Introduction to mixed methods in impact evaluation', *Impact Evaluation Notes* 3, August, The Rockefeller Foundation, New York.

Barr, J. and Christie, A. (2014) *Better Value for Money: An Organising Framework for Management and Measurement of VFM Indicators*, Itad, Hove.

Davidson, E.J. (2005) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*, Sage, Thousand Oaks, CA.

Davidson, E.J. (2014) 'Evaluative reasoning', *Methodological Briefs: Impact Evaluation 4*, UNICEF Office of Research, Florence [online]. Available at: https://www.unicef-irc.org/publications/749-evaluative-reasoning-methodological-briefs-impact-evaluation-no-4.html

DFID (2011a) 'DFID's Approach to Value for Money (VfM)', UK Department for International Development.

DFID (2011b) 'Reviewing and Scoring Projects. How-To Note. A DFID Practice Paper', November, UK Department for International Development.

DFID (2016a) 'Raising the Standard: The Multilateral Development Review 2016', UK Department for International Development.

DFID (2016b) 'Rising to the Challenge of Ending Poverty: The Bilateral Development Review 2016', UK Department for International Development.

DFID (2019) 'VfM Guidance: The 4th E Equity—External Version', UK Department for International Development.

DFID (2020) 'DFID's Approach to Value for Money—Guidance for External Partners', June, Finance and Performance Department, UK Department for International Development [online]. Available at: www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiCpeOv0sf7AhWA1jgGHZQaBrkQFnoECA8QAQ&url=https%3A%2F%2Fwww.lampdevelopment.org%2Fwp-content%2Fuploads%2F2020%2F12%2FSmart-Guide_-Approach-to-Value-for-Money_External.docx&usg=AOvVaw10RlaugIp_3STh3HpUjJIU.

Donaldson, S., Christie, C.A., and Mark, M. (2015) *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations. 2nd Edition*, Sage.

Donor Committee for Enterprise Development (2016) 'Measuring Results & the DCED Standard' [online], available at: www.enterprise-development.org/measuring-results-the-dced-standard/.

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., and Stoddard, G.L. (2005) *Methods for Economic Evaluation of Health Care Programs*, Oxford University Press, Oxford.

Eoyang, G. and Oakden, J. (2016) 'Adaptive evaluation: A synergy between complexity and evaluation practice', *Emergence: Complexity and Organization*, 31 December. doi: 10.emerg/10.17357.e5389f5715a734817dfbeaf25ab335e5

FCDO (n.d.) 'Private Sector Department: VfM Guidance for Technical Assistance', UK Foreign, Commonwealth and Development Office.

Fleming, F. (2013) 'Evaluation Methods for Assessing Value for Money' [online], available at: www.betterevaluation.org.

Fournier, D.M. (1995) 'Establishing evaluative conclusions: A distinction between general and working logic', in D.M. Fournier (ed.), *Reasoning in Evaluation: Inferential Links and Leaps. New Directions for Evaluation* 58, pp. 15–32.

FSD Africa (2022) 'Value for Money Framework: Second Edition', prepared for FSD Africa by Julian King & Associates Ltd (reproduced with permission).

Gargani, J. (2018) 'Reflections on Social Impact Measurement and Management', keynote address, the Social Impact Management Initiative (SIMI) Social Impact Day, Tokyo, Japan.

Greene, J. (2005) 'The generative potential of mixed methods inquiry', *Westminster Studies in Education* 28(2), pp. 207–211.

Hansford, F., Akkurt, O., King, J., Vanzetti, D., Rasulova, S., and Namukasa, E. (2019) 'Independent Evaluation of TradeMark East Africa Deliverable 6D: Strategy 1 Value for Money Assessment', Oxford Policy Management [online], available at: www.trademarkea.com/wp-content/uploads/dlm_uploads/2020/01/TMEA-Value-for-Money-Assessment-FINAL.pdf.

HM Treasury (2022) 'The Green Book: Central Government Guidance on Appraisal and Evaluation', Government of the United Kingdom [online], available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063330/Green_Book_2022.pdf.

ICAI (2011) 'ICAI's Approach to Effectiveness and Value for Money', November.

ICAI (2018) 'DFID's Approach to Value for Money in Programme and Portfolio Management: A Performance Review', February.

Itad (2012) 'Better Assessing VFM in DFID Nigeria Governance and Conflict Programming', October.

Jackson, P. (2012) 'Value for Money and International Development: Deconstructing Myths to Promote a More Constructive Discussion', OECD Development Co-operation Directorate.

King, J. (2015) 'Letter to the editor: Use of cost–benefit analysis in evaluation', *Evaluation Journal of Australasia* 15(3), September, pp. 37–41.

King, J. (2016) 'Value for Investment—A Practical Evaluation Theory' (e-book), available at: www.julianking.co.nz/wp-content/uploads/2014/09/160527-VFI-jk8-web.pdf.

King, J. (2017) 'Using Economic Methods Evaluatively', *American Journal of Evaluation* 38(1), March.

King, J. (2019a) 'Evaluation and Value for Money: Development of an Approach using Explicit Evaluative Reasoning', PhD, University of Melbourne [online], available at: https://minerva-access.unimelb.edu.au/handle/11343/225766).

King, J. (2019b) 'Combining multiple approaches to valuing in the MUVA female economic empowerment program', *Evaluation Journal of Australasia* 19(4), pp. 217–225 [online], available at: www.julianking.co.nz/wp-content/uploads/2019/12/King-2019-EJA-Praxis.pdf.

King, J. (2021) 'Expanding theory-based evaluation: Incorporating value creation in a theory of change', *Evaluation and Program Planning* 89, December [online], available at: https://www.julianking.co.nz/wp-content/uploads/2021/06/King-2021-ToVC-Preprint.pdf. https://doi.org/10.1016/j.evalprogplan.2021.101963

King, J. (2023a) 'Objectivity and Subjectivity in Evaluation' [online], available at: https://juliankingnz.substack.com/publish/posts/detail/106677217?referrer=%2Fpublish%2Fposts.

King, J. (2023b) 'Evaluation vs Value for Money: How VfM Assessment Differs (or not) from Evaluation' [online], available at: https://juliankingnz.substack.com/p/evaluation-vs-value-for-money.

King, J. (2023c) 'Rubrics and Alternatives' [online], available at: https://juliankingnz.substack.com/p/what-are-alternatives-to-rubrics.

King, J., McKegg, K., Oakden, J., and Wehipeihana, N. (2013) 'Rubrics: A method for surfacing values and improving the credibility of evaluation', *Journal of MultiDisciplinary Evaluation* 9(21), pp. 11–20.

King, J. and Guimaraes, L. (2016) 'Evaluating Value for Money in international development: The Ligada Female Economic Empowerment Program', *eVALUation Matters*, 3rd quarter, pp. 58–69, Africa Development Bank [online], available at: http://idev.afdb.org/sites/default/files/Evaluations/2020-02/Evaluation%20Matters%20Q3_EN.pdf.

King, J. and Allan, S. (2018) 'Applying evaluative thinking to Value for Money: The Pakistan Sub-National Governance Programme', *Evaluation Matters—He Take Tō Te Aromatawai* 4, pp. 207–235.

King, J. and OPM (2018) 'OPM's Approach to Assessing Value for Money: A Guide', Oxford Policy Management.

Levin, H. and McEwan, J. (2001) *Cost-effectiveness analysis (2nd ed),* Sage.

Mayne, J. (2008) Contribution Analysis: An approach to exploring cause and effect, *ILAC methodological brief.* Retrieved from: http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf

McKegg, K., Wehipeihana, N., Oakden, J., King, J. (2017) *Evaluative rubrics – delivering well-reasoned answer to real evaluative questions*. Presented at the Canadian Evaluation Society Conference, Vancouver, May 2017.

Mintzberg, H. and Waters, J. (1985) Of Strategies, Deliberate and Emergent. *Strategic Management Journal (6)*3, 257-272.

NAO (n.d.) 'Assessing Value for Money' [online], available at: www.nao.org.uk/successful-commissioning/general-principles/value-for-money/assessing-value-for-money/.

OECD (2021) 'Applying Evaluation Criteria Thoughtfully', Organization for Economic Co-Operation and Development.

Olson, E. and Eoyang, G. (2001) *Facilitating Organization Change: Lessons from complexity science (1st ed).* Wiley.

Patton, M.Q. and Campbell-Patton, C.E. (2021) *Utilisation-Focused Evaluation (5th ed)*, Sage.

Renard, R. and Lister, S. (2015) *Measuring and Reporting on Value for Money: A conceptual framework for MDBs*. Oxford: Mokoro Ltd.

Rogers, P. (2014) 'Theory of Change'. *Methodological Briefs: Impact Evaluation 2*. Florence: UNICEF Office of Research.

Schwandt, T.A. (2015) *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*, Stanford Business Books, Stanford, CA.

Scriven (1991) *Evaluation Thesaurus*. Newbury Park, CA: Sage.

Scriven (1995) 'The Logic of Evaluation and Evaluation Practice'. In DM Fournier (ed.). *Special Issue: Reasoning in Evaluation: Inferential Links and Leaps*. *New Directions for Evaluation,* 68: 49–70.

SROI Network (2012). *Guide to Social Return on Investment.* The Social Return on Investment Network, January 2012.

UK Government Homes and Community Agency (2014). *Additionality Guide (4th ed).* https://www.gov.uk/government/publications/additionality-guide

Yarbrough, D.B., Shulha, L.M., Hopson, R.K., and Caruthers, F.A. (2011) *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.), Sage.

Annex A Economic analysis

Economic analysis can often provide part of the mixed methods evidence needed to support a VfM evaluation. It is beyond the scope of this guide to provide instruction on how to conduct economic analysis. For an introductory text on the subject, see Levin and McEwan (2001); for a more comprehensive text, see Drummond *et al.* (2005). The Green Book (HM Treasury, 2022) provides government guidance on the appraisal of public investments in the UK. The box below provides a snapshot of the main types of economic evaluation method.

Box 7: Economic methods of evaluation

'Economic evaluation offers a powerful set of methods for considering the costs and consequences of resource allocation decisions. All of these methods involve systematically identifying, measuring, valuing and comparing the costs and consequences of alternative courses of action (Drummond *et al*., 2005). All the methods yield indicators of efficiency but differ in their scope and the units of measurement used.' (King, 2017: 103) The methods typically consider a stream of costs and consequences over a defined time horizon, with discounting to adjust the values of costs and consequences according to when they occur in time (with present costs and consequences being valued more highly than future costs and consequences).

'Cost-effectiveness analysis (CEA) measures costs in monetary terms, and consequences in natural or physical units—usually, a single quantifiable outcome measure with a strong counterfactual, such as years of life saved by a health intervention. The output of a CEA is a *cost-effectiveness ratio* (e.g. the average cost per year of life gained). Often, an *incremental cost-effectiveness ratio* is calculated that compares the *additional* costs and consequences of an intervention to those of its next-best alternative.' (King, 2017: 103)

'Cost–utility analysis is closely related to CEA in terms of its underlying structure but broadens the valuation of consequences to incorporate the notion of utility to people, which may include multiple attributes. For example, empirically-derived measures such as quality-adjusted life years scale the "raw" measurement of extended lifespans to take into account the utility of those additional years.' (King, 2017: 104) The 'incremental cost–utility ratio' is the average cost per unit of improvement in the utility index value compared to the next-best alternative.

'Cost–benefit analysis values all costs and consequences in the same units, which are usually monetary. In practice, some of the most valuable outcomes in a social programme can be difficult to value in monetary terms; however, in principle all values should be included and there are well-established economic valuation methods to support this approach' (King, 2017: 104). In practice, this often involves reference to market valuations to derive proxy values for use in the cost–benefit model (e.g. the value of training and subsequent employment in computer coding could be represented by the average market wage for coders and the probability of a trainee finding relevant employment). For intangible benefits, where no market exists (e.g. the value of friendship), economists can set up 'hypothetical markets' in the form of experiments to elicit approximate monetary values. Approaches to monetising wellbeing based on subjective measures such as life satisfaction are also coming into use (HM Treasury, 2022). The output of a cost–benefit analysis can take various forms such as net present value (benefits minus costs), benefit–cost ratio (benefits divided by costs), or return on investment (net present value divided by costs). (King, 2017)

Although there is no hard-and-fast rule to say when economic methods are fit for purpose, there are some minimum requirements. These include having access to adequate data, specialist skills, and sufficient resources to conduct a sound economic analysis.

Economic evaluation can enhance evaluation of VfM in several ways.

- 'Systematically evaluating costs and consequences yields insights that cannot be gained by looking at either factor in isolation' (King, 2017: 104). For example, two alternative interventions may be equally effective but differ in regard to their costs.
- 'Economic evaluation often involves modelling (creating a simplified representation of a system to facilitate analysis and understanding) and forecasting (estimation of future value) as well as measurement of past performance. Both modelling and forecasting introduce assumptions and uncertainty. A strength of economic evaluation is the ability to apply scenario analysis (exploring the value of a programme under different sets of assumptions or circumstances), sensitivity analysis (exploring the extent to which changes in a particular input variable affect the outputs of the model)' (King, 2017: 104), and break-even analysis (identifying the conditions or assumptions under which benefits would equal costs). This facilitates transparency and robust thinking about costs and consequences.
- Economic analysis communicates value in a language that is familiar and credible to donors and other decision makers, and can help shift the focus from the 'cost' of a programme onto the 'value of the investment' (King, 2015).

However, economic evaluation on its own is usually insufficient to support a comprehensive evaluation of VfM. Economic evaluation principally estimates economic efficiency, or whether overall welfare at a societal level increases or decreases. This is often an important consideration in VfM, but it needs to be balanced with other considerations, such as affordability, relevance, equity, and sustainability. Moreover, in practice, economic evaluation may provide an incomplete, or even misleading, picture of economic efficiency because it is often too difficult to include robust measurements of some of a programme's benefits and costs (King, 2017).

Within a VfM evaluation, the output of an economic evaluation—e.g. an incremental cost-effectiveness ratio, cost–utility ratio, benefit–cost ratio, or net present value—can be treated as an indicator that feeds into the wider evaluation of VfM (King, 2017; 2019).