OPM's approach to assessing Value for Money

Report

January 2018

Authors

Julian King

OPM VfM Working Group

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) and Director of Real Evaluation LLC, Seattle

Foreword

As aid budgets come under increasing scrutiny, aid agencies and development practitioners face a growing need to demonstrate value for money (VfM). This is appropriate and should be welcomed. However, up until now this need had not been matched with appropriate methods to support meaningful VfM assessment, with the prevailing tendency to rely on readily quantifiable measures, even when those don't capture the most important aspects of the change that is 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 the basis of a robust and distinct OPM approach to assessing VfM. The methodology was successfully piloted during the annual reviews of the Department for International Development's (DFID) Sub-National Governance programme in Pakistan and MUVA, a women's economic empowerment programme in Mozambique. The approach involves making transparent, evidence-based judgements about how well resources are being used, and whether the value derived is good enough to justify the investment.

To date, we have applied this approach on upwards of a dozen different 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 in particular appreciate the use of explicit evaluative reasoning. This, involves developing definitions of what acceptable / good / excellent VfM looks like, in the context of each specific project. Critically, these definitions are co-developed and endorsed upfront, in advance of implementation and before the evidence is gathered, which provides an agreed, objective, and transparent basis for making judgements.

Responding to high demand, OPM's VfM working group has developed this short and practical guide on our approach to VfM, which details the underlying conceptual framework and sets out the practical steps required for applying it, including how these can be integrated with routine project management activities. It is our hope that this will make a valuable contribution to strengthening accountability and impact of development interventions.

Simon Hunt, Managing Director, OPM

Table of contents

Acknowledgements

The authors are grateful for the input of OPM staff into this document through VfM workshops in Oxford (May 2017) and through the development of the MUVA and Sub-National Governance Programme (SNG) VfM frameworks during 2016 – in particular, Stephanie Allan, David Hoole, Luize Guimaraes, Martin Rimmer, Henlo Van Nieuwenhuyzen, Alex Hurrell, Nils Riemenschneider, Alex Jones, Jenny Congrave, Lucie Moore, Daniel Wate, Astrid Gronbaek, Tracey Johnson, Martina Garcia Aisa, and Stephen Akroyd. This document draws on a practice-based body of knowledge about evaluative reasoning developed through 10 years' collaborative work with Julian King's Kinnect Group colleagues Judy Oakden, Nan Wehipeihana, and Kate McKegg.

This document should be referenced as King and OPM (2018). *The OPM approach to assessing value for money: A guide.* Oxford: Oxford Policy Management Ltd.

Executive summary

This document offers practical guidance for assessing value for money (VfM) in international development programmes. VfM is an evaluative question about how well resources are used, and whether the resource use is justified (King, 2017). Addressing an evaluative question requires more than just measurement: it requires judgements to be made from the evidence. Good VfM assessment is clear about the basis upon which the judgements are made.

In international development there is increasing scrutiny on VfM but a lack of appropriate methods to support its assessment. There is a risk of reaching invalid conclusions if VfM assessment 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 (King and Guimaraes, 2016).

The approach presented in this guide combines cutting-edge evaluation practice with concepts from economic evaluation to respond to donor 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 and evaluation (M&E) systems – both for efficiency's sake, and to ensure conceptual coherence between VfM assessment and wider M&E work.

The VfM framework achieves these aims by:

- Using explicit criteria (dimensions of VfM) 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 the 'Four E's' approach to VfM assessment which is familiar to international aid donors.

The box overleaf briefly describes the application of this approach to assessing VfM on the MUVA female economic empowerment project in Mozambique.

Box 1: Measuring VfM on MUVA, a female economic empowerment programme

This framework shows how we can use the Four E's to support clear thinking. Just as a Theory of Change is an investment in reaching an explicit, shared understanding of the intent of the programme, we also invested time during the inception phase in defining what good economy, efficiency, effectiveness, cost-effectiveness and equity would look like in the MUVA programme. This investment proved most worthwhile when the time came to assess VfM, because we had an agreed and transparent basis for making robust judgements, and a clear framework for reporting VfM.

It was also really valuable that the framework incorporates economic analysis at the cost-effectiveness level. There is real value in understanding cost structures and providing hard data demonstrating, for example, that if we want quality it's not always cheap, if we want sustainable systems it takes time, and if we want to work with disadvantaged populations we may need to invest more than with highly educated ones. (Luize Guimaraes, MUVA Programme Manager, OPM)

MUVA is a six-year DFID-funded programme, led by OPM, which aims to link girls and women more closely with Mozambique's economic growth.[1] Despite impressive growth in recent years, over half of Mozambicans still live in poverty (OPM, 2015). Households headed by women are more likely to be among this number, with women's exclusion from decent economic opportunities a major cause: 19–24-year-old females in Mozambique's cities suffer the highest unemployment rates in the country (ibid.). It is this group of young, urban, and largely economically excluded women who are the focus of the MUVA programme, which aims to find innovative solutions to improving their recruitment and retention in secure, well-paid jobs and providing them with better access to markets.

Driven in part by its overarching objective – finding out how to achieve female economic empowerment in Mozambique through testing and learning – the programme has a unique structure and design. A long inception phase was included in the design to allow for decision making about interventions to be informed by studies to find out what would work best in the local context. The programme designed a system by which it would fund different project interventions in a selected number of identified female economic empowerment pathways.

How does one assess VfM for an innovative, learning programme?

VfM in MUVA is challenging for several reasons:

- MUVA is a learning programme. Some of the projects and approaches being tested have not been tried before, while others build on an existing evidence base. Part of the value of MUVA is derived from learning – including providing evidence about what works, what does not work, and why.
- MUVA is an adaptive programme. Its value may be enhanced by the iterative approach to testing, reflection and adaptation that helps to incrementally improve interventions – but this also means traditional approaches to measuring and valuing success (which assume constancy of an approach for a period of time) may not provide a valid assessment of VfM.
- MUVA is an influencing programme. One of its objectives is to influence stakeholders and social norms. The nature and full extent of these impacts will not become apparent until some time after the life of the programme – but it will be important to understand the direction of travel and associated value.
- Female economic empowerment brings tangible and intangible benefits. The intangible value of MUVA's outcomes (e.g. quality of life, agency, self-worth) are inherently valuable outcomes

[1] www.muvamoz.co.mz

for the target groups but are difficult to value credibly in monetary terms (King and Guimaraes, 2016).

An evaluation-specific approach to VfM

The MUVA VfM framework was developed using the approach described in this guide.

The steps in this approach included:

1. Establishing criteria of merit or worth – selected dimensions of performance relevant to VfM. The MUVA VfM framework built on and extended DFID's Four E's (DFID, 2011a) by developing explicit definitions for economy, efficiency, effectiveness, cost-effectiveness, and equity, tailored to the MUVA context. For example, criteria for effectiveness included three sub-dimensions (effectiveness as an urban female economic empowerment programme, effectiveness as a learning programme, and effectiveness as an influencing programme) – with detailed definitions and sub-criteria developed for each of these.

2. Defining performance standards – levels of VfM (i.e. definitions of excellent, good, adequate, and poor performance).

3. Gathering and analysing evidence of performance – including a mix of quantitative and qualitative evidence – to support a nuanced understanding and sound judgements of performance and VfM, as well as additional evidence aimed at contributing to the body of knowledge about the cost and value of investing in a programme of this nature.

4. Synthesising the results into an overall judgement, and preparing a clear, compelling performance story at regular intervals, tracking the performance and VfM of MUVA over time.

1 Background

This document offers practical guidance for carrying out VfM analysis in different contexts, primarily to meet the requirements of DFID but with wider applicability to other development partners. The approach presented in this guide combines good evaluation practice with concepts from economic evaluation, to meet DFID's requirements for VfM analysis (DFID, 2011a), 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 data (economic and other indicators) and qualitative narrative.

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

1.1 What is VfM?

In general terms, VfM can be regarded an **evaluative question** about how well resources are used, and whether they are being used well enough (King, 2017). 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 – though numeric evidence plays an important role in supporting well-informed judgements.

Although there is no universal definition of VfM (King, 2017), 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 last of these. For example, DFID defines the VfM of its programme spend in terms of 'maximising the impact of each pound spent to improve poor people's lives' (DFID, 2011a: 2).

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 in 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, and meeting 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 assessment can nonetheless evaluate 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-term and medium-to-long-term outcomes, and agreed definitions of what 'good' performance would look like – informed by comparative data where possible.

VfM has political and bureaucratic drivers that stem from the need for accountability and transparency in the use of public resources. These drivers are evident in DFID publications such as the Multilateral Development Review (2016a), which mentions VfM 21 times and suggests that

DFID will increasingly back organisations that deliver and demonstrate good VfM. 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 15 years (Fleming, 2013).

Good VfM assessment can also support organisations to work effectively by helping to clarify and communicate the value of their work, and by providing insights to support learning and improvement. DFID's Bilateral Review (2016b) notes that improving VfM means not just measuring VfM and being accountable for spending and results, but also having systems for learning, adaptation, and improvement, to improve VfM at project, organisation, and system levels.

Although there is increasing scrutiny on VfM in international development, there has been a lack of appropriate methods to support VfM assessment. The objectives of this guide are therefore to assist donors and partner organisations to:

- Conduct hard-nosed assessments 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 assessment such as DFID's Approach to Value for Money (DFID, 2011a) and the Independent Commission for Aid Impact Approach to Effectiveness and Value for Money (Independent Commission for Aid Impact (ICAI), 2011), and Itad's Better Value for Money – an organising framework for management and measurement of VfM indicators (Barr and Christie, 2014) – among others.

The existing VfM guides are useful in setting out high-level principles for VfM assessment but fall short of providing a blueprint approach for assessing VfM. In particular, although both DFID's and ICAI's documents acknowledge that VfM involves **making judgements** from the 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 facilitate the use of mixed methods evidence to support better-informed judgements, including integrating economic analysis, as well as broader quantitative and qualitative evidence, into VfM assessments where appropriate and feasible.

1.4 Overview of this guide

Chapter 2 sets out the **conceptual framework** for the VfM approach. It highlights key concepts from evaluation theory as well as the need for the VfM approach to address DFID's VfM criteria and to integrate this with wider M&E frameworks.

Chapter 3 sets out the **practical approach** to designing, undertaking, and reporting a VfM assessment. 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.

2 Conceptual framework for VfM assessment

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

- Explicit evaluative reasoning, using criteria and standards ('rubrics'), provides a framework for making sound, traceable judgements about VfM.
- Although DFID's Four E's do not necessarily represent the full set of VfM criteria, they provide a reasonable starting point for thinking about VfM that will meet DFID's and many other donors' requirements.
- Programme-specific definitions of economy, efficiency, effectiveness, cost-effectiveness, and equity need to be developed, together with standards describing different levels of performance for each criterion – and agreed in advance of the VfM assessment.
- Evidence will usually come from a mix of methods including existing M&E data (quantitative and qualitative), and ideally economic analysis, supplemented with additional evidence according to needs.
- The VfM assessment process should be integrated with relevant M&E frameworks and processes – both for efficiency's sake, and to ensure conceptual alignment between VfM and related M&E work.
- VfM is a matter of context and perspective; for example, the VfM of a programme from the perspectives of beneficiaries and in-country governments may diverge from its VfM from the donor's perspective.

2.1 Explicit 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. These definitions are developed before the evidence is gathered, providing an agreed basis for making judgements.

Although there is more than one way to approach this task (e.g. see Schwandt, 2015; King, 2017), a widely used approach is to:

1. Establish criteria of merit, worth or significance – the dimensions of performance that are relevant and important to an evaluative judgement;

2. Define performance standards for each criterion, in order to distinguish the difference 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 assessments, but is a prime example of 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 Chapter 3. For further reading on evaluative reasoning, see Davidson, 2014.[2]

2.2 VfM criteria

VfM criteria are selected dimensions of programme performance that are relevant to the overall determination of the merit, worth, and significance of resource use. They describe at a broad level the aspects of performance that need to be evidenced to support an evaluative judgement about VfM. We can conceptualise these criteria as falling into three key groups:

- The resource use ('what did we put in, and how was it used?') – e.g. relevance, affordability, ethicality, fidelity, input efficiency, technical efficiency, allocative efficiency, adaptive management;
- Value derived from the resource use ('what did we get out?') – e.g. outputs, outcomes, impacts and their value; and
- Whether the value derived justifies the resource use ('was it worth it?') – e.g. outcome efficiency, equity, sustainability (King, 2016).[3]

Although not exhaustive (no official list of VfM criteria exists), this list shows that the 'merit, worth and significance of resource use' (King, 2017: 110) is a multi-dimensional concept and encompasses all of the OECD Development Assistance Committee (DAC) Criteria for Evaluating Development Assistance, as well as additional factors. This does not mean we have to use all of these criteria every time – but it is worth bearing in mind that VfM is a broad concept and any of these criteria could apply.

Although VfM and evaluation are conceptually related, pragmatic decisions need to be made about what is within the scope of a VfM assessment and what is part of the remit of a wider evaluation. DFID, for example, requires VfM assessments on a different (usually more frequent) cycle than summative evaluations, and the VfM assessment can be largely limited to the Four E's (described in the next section), with judicious inclusion of other criteria where they are considered so critical to VfM that their exclusion would result in poor judgements being made.

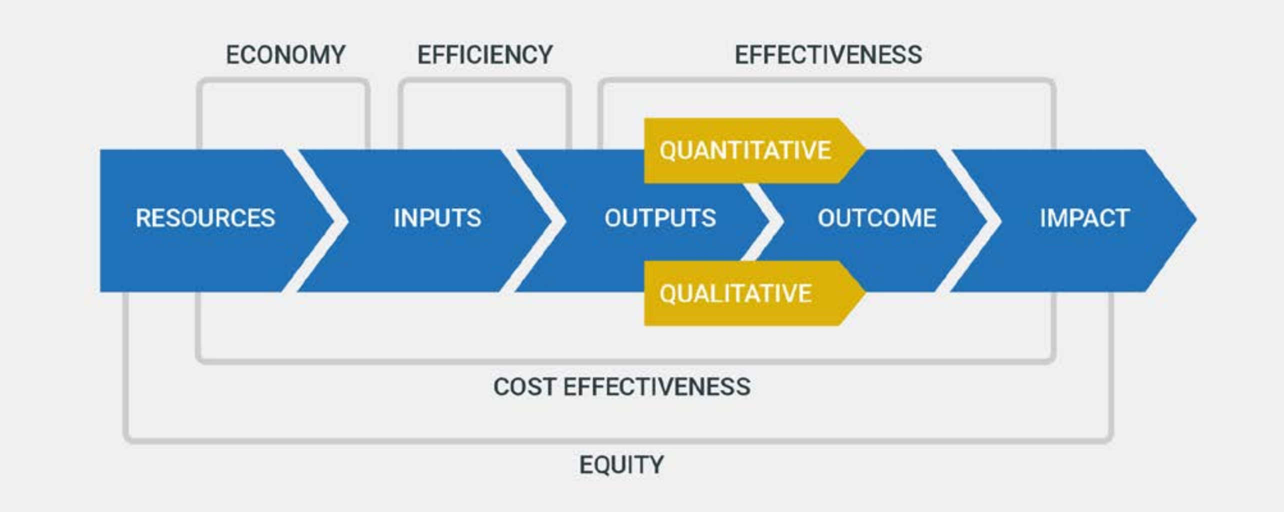
2.3 DFID's VfM criteria: the Four E's

DFID are thought leaders in the assessment of VfM in development programmes, having funded various studies on the subject (see, for example, Itad 2012). It has also been a subject of interest for the Independent Commission on Aid Impact, which scrutinises UK aid spending (see, for instance, ICAI 2011). When evaluating VfM in DFID-funded programmes, the criteria to be used are specified in DFID's Approach to Value for Money (DFID, 2011a). This framework focuses on the 'Three E's' of economy, efficiency, and effectiveness. These levels correspond respectively to

[2] www.unicef-irc.org/publications/pdf/brief_4_evaluativereasoning_eng.pdf
[3] Opportunity cost is relevant to some of these dimensions but not to others. For example, efficiency is a comparative concept and economic evaluation explicitly considers the opportunity cost of an intervention by comparing it to the next-best alternative use of resources. Some other dimensions of VfM (e.g. affordability and ethicality) are not comparative; we want to know whether the investment is affordable and ethical in its own right. Some other dimensions are more ambiguous (e.g. equity, sustainabiity) and it will be up to the evaluator to decide whether it is appropriate and feasible to evaluate them comparatively or in an absolute sense.

three different levels of a programme results chain – inputs, outputs, and outcomes. A fourth 'E' – equity – is also acknowledged in the framework. Although the framework is commonly referred to as the Four E's, **there are actually five criteria**, the additional criterion being cost-effectiveness, which compares outcomes or impact to total costs.

Figure 1: VfM conceptual framework (DFID, 2011a)



DFID (2011a: 4) defines these criteria as follows.

> *Economy: Are we or our agents buying inputs of the appropriate quality at the right price? (Inputs are things such as staff, consultants, raw materials and capital that are used to produce outputs)*
>
> *Efficiency: How well do we or our agents convert inputs into outputs? (Outputs are results delivered by us or our agents to an external party. We or our agents exercise strong control over the quality and quantity of outputs)*
>
> *Effectiveness: How well are the outputs from an intervention achieving the desired outcome on poverty reduction? (Note that in contrast to outputs, we or our agents do not exercise direct control over outcomes)*
>
> *Cost-effectiveness: How much impact on poverty reduction does an intervention achieve relative to the inputs that we or our agents invest in it?*

DFID does not define equity explicitly, but does acknowledge the importance of distributive fairness as a dimension of VfM:

> *When we make judgements on the effectiveness of an intervention, we need to consider issues of equity. This includes making sure our development results are targeted at the poorest and include sufficient targeting of women and girls.* (DFID, 2011a: 3)

The most important measures of VfM relate to cost-effectiveness and equity. Performance on economy, efficiency, and effectiveness can provide important diagnostic information, help identify where inefficiencies may be occurring, and contribute to the overall assessment.

DFID's 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 having been set in the appraisal phase, procurement emphasises selection of service providers who can deliver economically and efficiently. Programme implementation aims to balance all Four

E's, though it is reasonable to be more reliant on process indicators (rather than results) to assess VfM in the early stages of a programme.

The inclusion of equity in the framework is particularly important because it indicates that the VfM of development aid depends not just on efficient delivery and cost-effective outcomes, but on reaching women and girls, young people, poor people, people with disabilities, people living in more remote rural areas, and other marginalised groups – and delivering improvements in equity as efficiently and cost-effectively as possible. Achieving this involves a balancing act – e.g. 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 the trade-offs between them.

2.4 Limitations of the Four E's

The Three E's of economy, efficiency and effectiveness divide the concept of VfM into discrete sections of a programme's theory of change. **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, this simplified depiction of a results chain 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 learning).

The simplified results chain also omits some steps that may be important for VfM. For example, in between inputs and outputs, a full theory of change 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 judgements about whether a programme represents VfM.[4] Statements on VfM which rely on indicators at lower levels, as may be the case in the early stages of programme implementation when outputs and outcomes have not yet been delivered, need to be duly qualified.

The terminologies used in DFID'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 DFID restricts this term to the output level alone (Renard and Lister, 2015). Even at output level, DFID's definition of efficiency seems restricted to the concept of 'technical efficiency' (maximising outputs for a given level of inputs), whereas VfM assessment 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).

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

Moreover, in complex adaptive programmes, 'responsiveness to context' may be more important for VfM than static concepts of how efficiently inputs are converted to 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), which measures the ratio between a single outcome indicator, measured in natural or physical units (e.g. lives saved), and monetary programme costs. In VfM assessment, a broader interpretation is needed: the 'cost-effectiveness' criterion may be better labelled 'outcome efficiency' and should address whether the programme creates more value than it consumes, or whether sufficient outcomes or impacts are achieved to justify the investment of resources. 'Value created' potentially includes both tangible (monetary or readily monetisable) and intangible value (outcomes or impacts that are difficult to value credibly in monetary terms).

2.5 Defining criteria and standards for the Four E's

The Four E's 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 the concepts directly relatable and meaningful to the programme at hand.

Example: MUVA effectiveness criteria

In the MUVA female economic empowerment programme in Mozambique, **effectiveness** was broken down into three sub-criteria, which were defined as follows (King and Guimaraes, 2016):

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.

As can be seen from this example, 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, and these descriptions are deliberately broader and less specific, reflecting their purpose of facilitating meaningful evaluative judgements.

Programme-specific definitions of the Four E's are useful but do not 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) provides a traffic light rating system which represents a generic set of standards to support judgements about effectiveness and VfM (Table 1).

Table 1: Generic rubric (ICAI, 2011)

Rating	What it means
G	**Green:** The programme meets all or almost all of the criteria for effectiveness and value for money and is performing strongly. Very few or no improvements are needed.
G A	**Green-Amber:** The programme meets most of the criteria for effectiveness and value for money and is performing well. Some improvements should be made.
A R	**Amber-Red:** The programme meets some of the criteria for effectiveness and value for money but is not performing well. Significant improvements should be made.
R	**Red:** The programme meets few of the criteria for effectiveness and value for money. It is performing poorly. Immediate and major changes need to be made.

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 assist in making evaluative judgements. Programme-specific standards contain descriptions of relevant sub-criteria describing what performance looks like at each level. In so doing, such standards also indicate the relative importance of different criteria.

For example, the following performance standards for 'economy' were developed for FOSTER 2, the second phase of a project focused on helping Nigeria make more use of its extractive industry to support national development (Table 2).

Table 2: FOSTER 2 economy standards

Performance	Criteria
Excellent	Unit costs for national/international consultants and flights *consistently below* agreed benchmarks. Meets all criteria under 'good' performance.
Good	Unit costs for national/international consultants and flights *generally*[5] *at or below* agreed benchmarks. FOSTER 2 can demonstrate results of good cost management, e.g. partner contributions (monetary, pro-bono, and resources in-kind) and savings secured through contract negotiation/management. Meets all criteria under 'adequate' performance.
Adequate	Unit costs for national/international consultants and flights are *generally at or near* agreed benchmarks. Overheads as a percentage of total costs generally at or below agreed benchmark. FOSTER 2 verifiably follows good practices to manage the key economy drivers of: consultant selection, recruitment, and fee setting; procurement and benchmarking; administration and overheads management; fiduciary risk management; and economies of scale.[6]
Poor	Any of the conditions for 'adequate' not met.

[5] For the purposes of these criteria, 'generally' means for the most part, allowing for a few reasonable exceptions. As the purpose of the criteria is to facilitate a transparent and defensible judgement, it would be inappropriate to specify exact thresholds because such thresholds would be arbitrary and would not support meaningful assessment. All judgements made will be transparent and therefore challengeable.
[6] Detailed definitions were also developed for each of these economy drivers.

2.6 Mixed methods evidence

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 described 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.*

Examining the effectiveness criteria for MUVA, it is obvious that some of the evidence needed to support evaluative judgements is qualitative. Indicator-based measurement makes a valuable contribution to evaluating programme performance and VfM. Indicators alone, however, may be 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 correlate with the VfM criteria. Broader contextual evidence is also important to provide further information about performance and support appropriate interpretation of the indicators.

To support sound VfM assessment, we need to draw on a balance of quantitative measures and combine this with the types of information that come from having real conversations with real people. A mix of evidence (numbers and stories) can give us a better understanding of whether a programme is really changing lives.

Five key reasons for using mixed method designs include:

- **Triangulation of evaluation findings** – enhancing the validity or 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 (Greene, 2005).

2.7 Economic analysis

Economic analysis can often provide part of the mixed methods evidence needed to support a VfM assessment. 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 box below provides a snapshot of the main types of economic evaluation methods.

Box 2: 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 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. The output of a cost–benefit analysis can take various forms such as net value (benefits minus costs), benefit–cost ratio (benefits divided by costs), or return on investment (net 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) and *sensitivity analysis* (exploring the extent to which changes in a particular input variable affect the outputs of the model)' (King, 2017:104). 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 to the 'value of the investment' (King, 2015).

However, economic evaluation on its own is usually insufficient to support a comprehensive assessment of VfM. Economic evaluation principally estimates efficiency – and efficiency is only one consideration, which has to be balanced against others such as affordability, relevance, equity, sustainability, etc. Moreover, in practice, economic evaluation may provide an incomplete picture of 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 assessment, 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).

2.8 Complexity and 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.

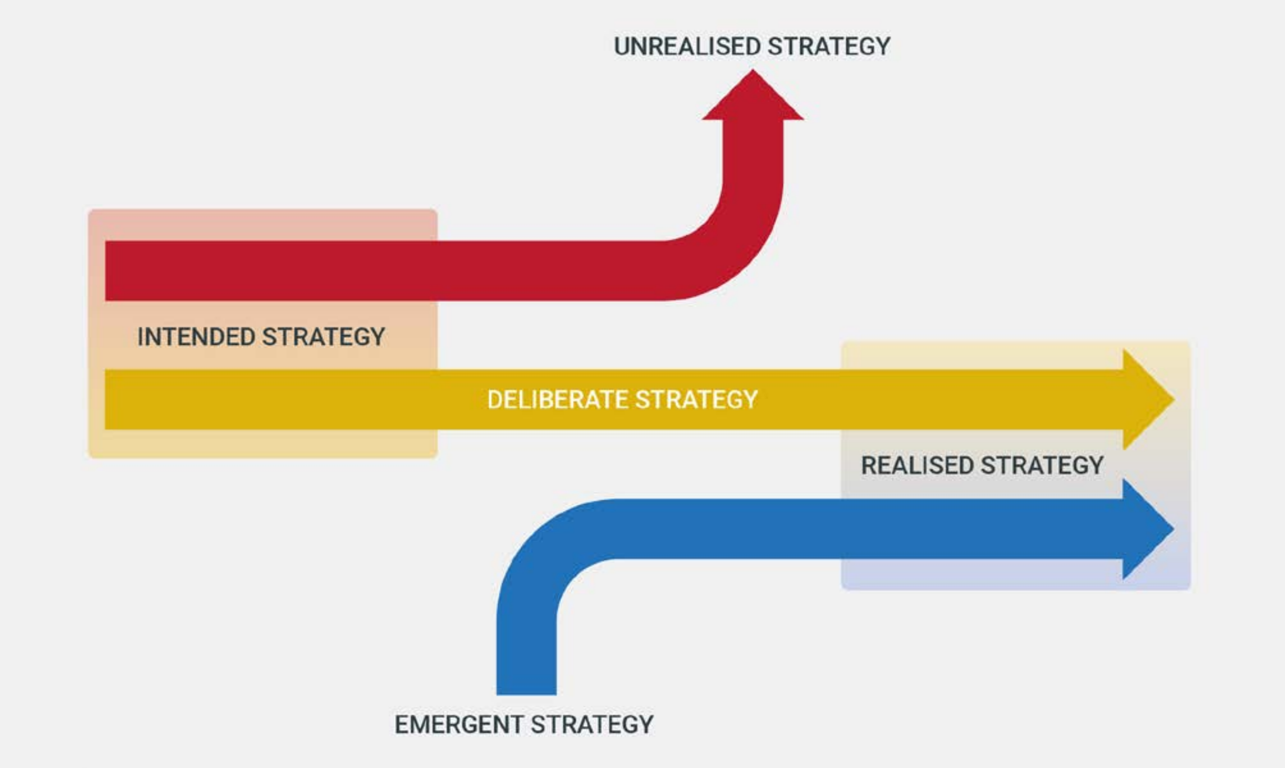
There is a body of literature on how complexity approaches differ from more linear management approaches.[7] For example:

- 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 inter-relationships (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).[8]

Accordingly, development programmes must be responsive to their evolving context. There will be some aspects of strategy that are planned ('intended strategy') but 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 below in Figure 2.

[7] For example, see Olson and Eoyang (2001).
[8] Judy Oakden (2017), personal communication.

Figure 2: Emergent strategy



Based on Mintzberg and Waters (1985).

The box below briefly describes how the VfM assessment of the SNG was able to take account and reflect emerging strategies.

Box 3: The VfM assessment of SNG – a complex governance programme

SNG is supporting local governments in two provinces in Pakistan (Khyber Pakhtunkhwa and Punjab) to improve the delivery of basic services. Financed by DFID, and managed by OPM, the programme operates 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 is that 'poor people in Punjab and Khyber Pakhtunkhwa report that services are better meeting their needs', and its intended impact is 'a more stable democracy in Pakistan, through increased trust in Government'.

SNG is a complex governance reform programme. Its complexity is derived in part from its operating environment, which necessitates 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 central and district governments during a period of considerable decentralisation reform. The VfM assessment needed to be cognisant of this context, and assess the programme's relevance in a complex and rapidly changing environment.

An additional challenge in assessing VfM of SNG is that the value of the programme lies not only in the achievement of the specified outcomes, but also critically in 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 is inherent to the 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, hence 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.

2.9 Perspective matters in VfM

> *'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.

For example, a DFID-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 make good VfM assessments 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 assessment should seek to empower the voices of 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 this 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).

2.10 Integration with M&E frameworks

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

Not all M&E tasks involve a VfM assessment. Nevertheless, a VfM assessment should align as much as possible with wider M&E work for two important reasons:

- Conceptually, the VfM and M&E frameworks for a programme should be coherent with each other; for example, they should be based on the same theory of change and should use criteria and indicators that are consistent (or at least not contradictory).
- Practically, the VfM and M&E frameworks for a programme 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.

Given the diverse demands of evaluation clients and plurality of evaluation approaches, the VfM approach needs to be adaptable so that it can be integrated with the evaluation context of each programme.

Chapter 3 identifies the points along the VfM process where the VfM framework design and reporting should be integrated with other M&E activity.

2.11 Timing of VfM assessments

This VfM approach can be used to support regular VfM assessments and therefore to monitor and track VfM over time. The frequency of VfM assessment is flexible and should take into account the nature of a programme's intended outcomes and the expected time frames to achieve substantive outcomes. For example, it might make sense to track some outcomes on an annual basis, whereas for longer-term impacts that take longer to achieve, assessments 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.

2.12 Level of effort required

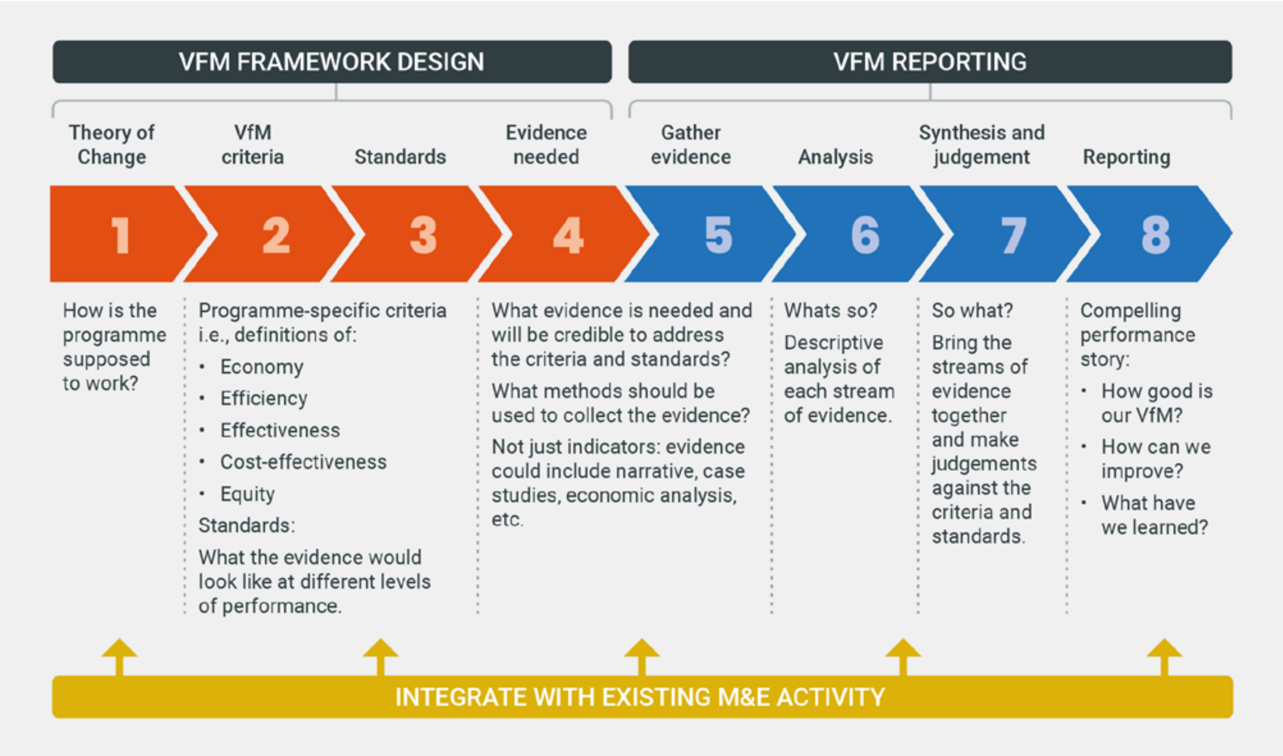
A full VfM assessment using this framework can be a time-intensive undertaking, and consideration must be given to proportionality when assessing the VfM of smaller programmes. The use of criteria and standards to guide evaluative judgements must be tailored to the scope and scale of the programmes,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

3 Designing and implementing a VfM framework

This chapter sets out a practical approach for designing, undertaking, and reporting a VfM assessment. 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 M&E activity are identified.

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

Figure 3: Overview of VfM approach[9]



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 assessment 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 specifying a theory of change, identifying VfM criteria, developing standards for each criterion, and determining the evidence necessary to support well-informed evaluative judgements against the criteria and standards.

3.1 Step 1: theory of change

A theory of change '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

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

theory of change 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 theory of change before moving on to the development of criteria and standards and then 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 theory of change, as this is well covered in other texts (e.g. Rogers, 2014). Both the VfM assessment and any other M&E activity should use the same theory of change, so that the evaluation frameworks are consistent and coherent across all M&E activity. This may affect decisions about when to develop the theory of change, and who should be involved (e.g. M&E staff as well as the person who will be in charge of VfM assessment, together with any other stakeholders who would normally be involved).

Typically, a theory of change 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 theory of change to ascertain whether it is fit for purpose, adding to or amending it as necessary.

Example: Identifying SNG's intermediary outcomes

In the SNG programme, the theory of change 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 M&E advisers, technical advisers and management to review the theory of change at intermediary outcome level, unpacking the implied causal pathway towards outcomes, and validating the intermediary outcomes with the project teams.

Additionally, it was noted that the value of the SNG programme lay not only in the achievement of the specified outcomes, but critically in its capacity to generate learning that influenced other relevant programmes, and in its capacity to be adaptive, responding to lessons learnt as well as emergent opportunities and challenges as they arise, thus maximising the value of the DFID-funded technical assistance. These features were also therefore included in the VfM framework.

A good theory of change is explicit about resources and inputs and how they are supposed to be used to produce outputs and outcomes – so should already contain all of the information needed to support the identification of VfM criteria and standards.

3.2 Steps 2 and 3: VfM criteria and standards

As noted in the previous chapter, the performance of complex aid programmes should not be judged solely on the basis of indicators, devoid of any evaluative judgement. Well-reasoned judgements of VfM are required. 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 Four

E's as the basis for VfM assessment, criteria need to be defined for economy, efficiency, effectiveness, equity, and cost-effectiveness.

- **Performance standards** provide defined levels of VfM (i.e. excellent, good, adequate, and poor) for each of the criteria.

Criteria and standards should reflect the key inputs, activities, deliverables, outputs, intermediary and long-term outcomes, and impact articulated in the theory of change.

Developing VfM criteria and standards around the Four E's 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 sub-criteria that describe dimensions of good performance; and

3. Organise the sub-criteria into levels (e.g. definitions of excellent, good, adequate, 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 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 to the table to identify the values that form the basis of the criteria and standards. Although key requirements will be formally documented (e.g. in a contract, logframe or theory of change), the criteria and standards will elaborate on these requirements, which may involve reconciling or accommodating diverse perspectives (King *et al*., 2013).

Through this process, a shared understanding of what matters is reached, 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. p.14). Through their participation in this process, 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:

- 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 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 (King *et al*., 2013).

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) offers the following guidance:

> *Who should be involved in the evaluative reasoning process itself? Key considerations are:*
>
> - *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?*
> - *Cost – Taking into consideration the opportunity costs of taking staff away from their work to be involved in the evaluation, at which stages of the evaluative reasoning process is it best to involve different stakeholders for the greatest mutual benefit?(Davidson. 2014: 8)*

In practice, some of these considerations may be more important for a full summative evaluation of a programme than for a VfM assessment. 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, the time should be taken to ensure the donor fully understands and endorses 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 donors and other key stakeholders at the start.

Suspend conversations about measurement

During criteria development, it is important to 'park' any conversations about indicators, such as about how such indicators are to 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 in that they focus on the intended functioning and effects of the programme rather than 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:[10]

> *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 which define sector-specific criteria and standards, for example, the World Bank's International Benchmarking Network for Water and Sanitation Utilities; UNICEF's Child Well-Being Indicators; USAID's Health Systems Benchmarking Tool; the Public Expenditure and Financial Accountability (PEFA) Framework for public financial management, etc. 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.

Tailor to context

Good criteria are specifically tailored to the programme – and generally at a specified point in time (i.e. what would 'good' look like at the time of the VfM assessment). 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.

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, as far as possible. The following interpretation is suggested:

- **Excellent** means that the programme 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** means that the programme is generally meeting reasonable expectations/targets, allowing for a few minor exceptions. Some improvements may be needed.
- **Adequate** means that the programme, though not meeting all expectations/targets, is fulfilling minimum 'bottom-line' requirements and is showing acceptable progress overall. Significant improvements may be needed.
- **Poor** means that the programme is *not* fulfilling minimum 'bottom-line' requirements and/or *not* showing acceptable progress overall. Immediate and major improvements are needed.

[10] This is in line with a principle that is supported in economic evaluation: *Infeasibility of data collection does not mean irrelevance of data.* It is first important to work out what we need to know in order to inform a judgement, and then we need to work out whether we can get that information. Any information gaps are addressed in the discussion of uncertainty.

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 – e.g. we must not make unwarranted assumptions about programme results that are not knowable in advance.

Furthermore, criteria and standards need to be kept manageable. Instead of describing every possible dimension of a programme's intended performance, it is better to think in terms of the '80/20 rule' – i.e. what are the top 20% of possible criteria that would represent 80% of the important features of the programme's performance and VfM?

3.3 Step 4: identifying evidence required

In a logical and sequential process of evaluation design, it is only after clarifying the theory of change, criteria, and standards that relevant indicators and other 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. 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 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 of not just things we can count, or data we happen to have available to us, but what really matters and has some relevance to the VfM of the programme. Ideally this will include a mix of numeric, qualitative, and economic/monetary evidence.

Key sources of evidence to support the VfM assessment will vary with context but may, for example, include programme financial accounting data, log frame indicator data, programme operational data/reports, M&E data/reports, and results of economic analysis.

Prepare 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 theory of change 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 M&E framework); and
- A plan and timetable/cycle for conducting the VfM assessment(s).

3.4 Step 5: gathering evidence

Gathering the evidence needed for a VfM assessment involves following the same principles of good project management and fieldwork that would be followed for any evaluation or research project. Indeed, if the VfM framework is well aligned with wider M&E frameworks, the VfM assessment may be able to make use of relevant M&E 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? In some programmes there will be 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 strategies such as the Bradford Hill Criteria or another judgement-based approach (e.g. Davidson, 2005).

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

Example: Attribution on SNG

The results of the Pakistan SNG Programme are best conceptualised as the synergistic effects of interventions (e.g. ideas, influencing, training, resources, documents) produced by the SNG teams, and the adoption and dissemination of these and similar interventions by the provincial governments themselves. Additionally, other development programmes may contribute to the intended outcomes. In this context, and without the possibility of a measurable counterfactual, expressing the programme's contribution as a percentage is not particularly meaningful as it fails to recognise the synergies and interdependencies between these influences.

Nevertheless, such a percentage basis has in the past been required to meet DFID's expectations. Therefore, a framework was needed to facilitate transparent judgements about the contribution of the SNG programme to the outcomes observed. A checklist of economic considerations (deadweight, displacement, attribution, and drop-off), adapted from the Social Return on Investment Guide (Nicholls *et al*., 2012) was used to provide a systematic way of assessing the SNG programme's contribution to outcomes using a transparent rationale. Judgements were made for each output area individually and then for the programme overall, in a workshop with sector experts from the SNG teams.

Subsequent feedback from the DFID review team indicated that, given the uncertainties inherent in quantifying the contribution of a complex, multi-faceted programme with multiple external factors in play, the OPM team should not attempt to estimate a specific percentage contribution to the SNG programme. Instead, future VfM assessments should focus on explaining the ways in which the SNG programme contributed to the increase in health and education spending, and the quality of this spend.

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 done firstly for each VfM criterion individually (i.e. economy, efficiency, effectiveness, cost-effectiveness and equity); and secondly 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 straight average of the scores for each component; rather, **greater weight should be given to those criteria that are deemed more relevant at the time the VfM assessment is carried out**. For example, if good effectiveness and cost-effectiveness evidence is available, these criteria would generally be given greater weight than indicators focusing on inputs (e.g. economy) and outputs (e.g. efficiency), which receive more weight in the early stages of implementation.

If sufficient investment is made in developing sound criteria and standards, it is in the synthesis of findings that the investment pays off, because this provides the structure 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 assessment of a small programme, the sense-making and reporting steps need not be onerous – for example, the VfM assessment may consist of a few quantitative indicators and a two-hour workshop with key programme staff to discuss programme performance and make judgements against the criteria and standards with supporting rationale.

Who should make the judgements?

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

Our recommendation is that programme, M&E, donor, and independent perspectives are all important and will add value to a VfM assessment. 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 that can only come from those directly involved in the programme. A preferred forum is therefore 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, 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 sense-making. 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'). 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 anybody trained in an academic research discipline, 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: 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 key question to ask is where our 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 criticism confuses methods (how we gather evidence) with reasoning (how we make judgements). Quantitative methods are not as of right more objective or robust than qualitative methods.[11] Both are important, and if we privilege one over the other, we may miss important information. Either quantitative or qualitative data (or preferably, both together) can support a valid, traceable, and challengeable judgement – or, if poorly used and reported, an opaque and invalid one. The use of criteria and standards supports sound judgement-making by providing a transparent set of filters for interpreting the evidence.[12]

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 (the Four E's), addressing each 'E' 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 'E' sequentially, setting out a definition, relevant criteria and standards, a summary of evidence against them, and a judgement on performance.

Opportunities to improve VfM

The process of making judgements about VfM against the criteria and standards provides a systematic way to identify 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 pull the judgement down to 'good' or 'adequate' and would be identified as an area for improvement. The VfM report should identify these opportunities.

[11] For example, see Donaldson *et al*. (2015).
[12] For more on rubrics and perceived subjectivity, see: www.julianking.co.nz/blog/on-rubrics-and-subjectivity/

Bibliography

Adler, M.D. and Posner, E. A. (2006) *New foundations of cost–benefit analysis.* Cambridge, MA: Harvard University Press.

Bamberger, M. (2012) 'Introduction to Mixed Methods in Impact Evaluation'*. Impact Evaluation Notes*, No. 3, August 2012. New York: The Rockefeller Foundation.

Barr, J. and Christie, A. (2014) *Better Value for Money: An organising framework for management and measurement of VFM indicators*. Hove, UK: ITAD.

Davidson, E.J. (2005) *Evaluation Methodology Basics: The nuts and bolts of sound evaluation*. Thousand Oaks: Sage.

Davidson, E.J. (2014) 'Evaluative Reasoning'. *Methodological Briefs: Impact Evaluation 4*. Florence: UNICEF Office of Research. www.unicef-irc.org/publications/pdf/brief_4_evaluativereasoning_eng.pdf

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

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

DFID (2016a) *Raising the standard: the Multilateral Development Review 2016.* Department for International Development, United Kingdom.

DFID (2016b) *Rising to the challenge of ending poverty: the Bilateral Development Review 2016.* Department for International Development, United Kingdom.

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*. 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: Oxford University Press.

Fleming, F. (2013). *Evaluation methods for assessing value for money*. Retrieved from 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): 15–32.

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

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

ITAD (2012) *Better Assessing VFM in DFID Nigeria Governance and Conflict Programming*. October 2012.

Jackson, P. (2012) *Value for Money and International Development: Deconstructing myths to promote a more constructive discussion*. OECD Development Co-operation Directorate.

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. (2015) 'Letter to the editor: Use of cost–benefit analysis in evaluation'. *Evaluation Journal of Australasia*, Vol. 15, No. 3, September 2015: 37–41.

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

King, J., Allan, S. (2017) 'Evaluating Value for Money in International Development: Pakistan Sub-National Governance Programme' (article prepared for submission to academic governance journal).

King, J. and Guimaraes, L. (2016) 'Evaluating Value for Money in International Development: The Ligada Female Economic Empowerment Programme'. *eVALUation Matters,* Third Quarter, 2016. Africa Development Bank.

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*, Vol. 9 No. 21: 11–20.

Levin, H.M. and McEwan, P.J. (2001) *Cost-Effectiveness Analysis*. 2nd Edition. Thousand Oaks: Sage.

Mayne, J. (2008) *Contribution Analysis: An approach to exploring cause and effect*. ILAC methodological brief.

McKegg, K., Wehipeihana, N., Oakden, J., King, J. and Davidson, E.J. (2017) 'Evaluative Rubrics – Delivering well-reasoned answers to real evaluative questions'. Presented at *Canadian Evaluation Society Conference,* Vancouver, May 2017.

Mintzberg, H. and Waters, J.A. (1985) 'Of Strategies, Deliberate and Emergent'. *Strategic Management Journal,* Vol. 6, No. 3: 257–272.

Nicholls, J., Lawlor, E., Neitzert, E. and Goodspeed, T. (2012) *A Guide to Social Return on Investment*. January 2012. Haddington, United Kingdom: The SROI Network.

Olson, E.E. and Eoyang, G.H. (2001) *Facilitating Organization Change: Lessons from Complexity*. Wiley.

OPM (2015) *Pioneering programme to improve female economic empowerment in Mozambique. November 2015*. Oxford: Oxford Policy Management. Accessed: July 2016 at www.opml.co.uk/publications/news/pioneering-programmeimprove-female-economic-empowerment-mozambique

OPM (2016) *Sub-National Governance Programme: VfM framework*. Oxford: Oxford Policy Management.

OPM (2016) *MUVA Programme: VfM framework*. Oxford: Oxford Policy Management.

OPM (2017) *Sub-National Governance Programme: VfM Report*. Oxford: Oxford Policy Management.

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. (2015) *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Redwood City: Stanford University Press.

Schwandt, T., Ofir, Z., Lucks, D., El-Saddick, K. and D'Errico, S. (2016) 'Evaluation: a crucial ingredient for SDG success'. *Briefing: Governance, Policy and Planning*. April 2016.

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

Scriven, M. (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.

Scriven, M. (2012) 'The Logic of Valuing'. In G. Julnes (ed.). *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation. New Directions for Evaluation*, 133: 17–28.

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.). Thousand Oaks, CA: Sage

World Bank and OPM (2008) *The Political Economy of Policy Reform: Issues and implications for policy dialogue and development operations*. Social Development Department, World Bank. November 2008. Report No. 44288-GLB.

Contact us

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

www.opml.co.uk

David Hoole, Portfolio Lead,
Public Financial Management
(David.Hoole@opml.co.uk)

Stephanie Allan, Senior
Consultant, Public Financial
Management
(Stephanie.Allan@opml.co.uk)

Julian King & Associates
PO Box 41-339
St Lukes 1346
Auckland
New Zealand

www.julianking.co.nz

Julian King, Director

(jk@julianking.co.nz)