Oxford Policy Management

Value for Money (VfM) Framework for demand-responsive evidence services: Evidence Fund and K4DD

Final report

Priya Rampal, Arshdeep Kaur, Craig Bardsley, Patrick Ward and Julian King

October 2025

UK International Development
Partnership | Progress | Prosperity

JK&A

About Oxford Policy Management

Our vision is for fair public policy that benefits both people and the planet. Our purpose is to improve lives through sustainable policy change in low- and middle-income countries.

Through our global network of offices, we work in partnership with national stakeholders and decision makers to research, design, implement and evaluate impactful public policy. We work in all areas of economic and social policy and governance, including health, finance, education, climate change and public sector management. We have cross-cutting expertise in our dedicated teams of monitoring and evaluation, political economy analysis, statistics, and research methods specialists. We draw on our local and international sector experts to provide the very best evidence-based support.

Oxford Policy Management Limited

Registered in England: 3122495

Ground Floor

40-41 Park End Street

Oxford, OX1 1JD

United Kingdom

Tel: +44 (0) 1865 207 300

Fax: +44 (0) 1865 207 301

Email: admin@opml.co.uk

Website: www.opml.co.uk

Preface

This report has been produced by OPM, with feedback from FCDO. The team leader for this work was Patrick Ward and project manager was Priya Rampal. The team members were Craig Bardsley and Arshdeep Kaur. Julian King provided quality assurance. For further information, please contact Priya Rampal [priya.rampal@opml.co.uk]. The reference number for the project is A6115.

The study was supported by the Global Science Department (GSD) of the Foreign, Commonwealth & Development Office (FCDO), UK Government. The views expressed herein do not necessarily reflect the official policies of the UK Government. The contact point for FCDO is Ludhiya Johnson [ludhiya.johnson@fcdo.gov.uk].

We extend our sincere gratitude to the Foreign, Commonwealth & Development Office (FCDO) for their generous support and funding throughout this initiative. We would like to thank Ludhiya Johnson, Anirban Ganguly, and Harry Achillini for their continued guidance, strategic insights, commitment to advancing inclusive and impactful development programming and being members of the core stakeholder group.

We are equally grateful to the participants of our stakeholder workshops - James Georgalakis, Katherine Cane, and Mark Herbert and members of the core stakeholder group - whose thoughtful contributions and candid reflections enriched our understanding and helped shape the evaluative framework.

Our appreciation also goes to all individuals who participated in Key Informant Interviews (KIIs). Their time, perspectives, and openness were invaluable in grounding our analysis in lived realities and institutional nuance. We are particularly grateful to Emmeline Skinner, Georgia Watson, and Samuel Bayes for providing feedback on the framework in an additional workshop which we conducted with them.

This work would not have been possible without the collective engagement and generosity of everyone involved. Any errors and shortcomings remain entirely our own.

Executive summary

This document presents a Value for Money (VfM) framework for the Evidence Fund (EF) and the Knowledge for Development and Diplomacy (K4DD) programmes. It was developed by Oxford Policy Management. These FCDO programmes are designed to provide timely, high-quality, and contextually relevant research and learning services that inform FCDO's policy, programming, strategic, and diplomatic decisions. The VfM framework offers a systematic approach to assessing programme performance across five dimensions - Economy, Efficiency, Effectiveness, Equity, and Cost-Effectiveness. It is hoped that it will also be useful, with appropriate adaptation, for similar demand-responsive research and evidence initiatives.

The Evidence Fund is a consolidated £58 million programme that commissions primary and secondary research, evaluations, and evidence syntheses to support strategic decision-making across FCDO and HMG. It streamlines previously fragmented research platforms to enhance coherence, efficiency, and learning. Research requests are usually assessed through structured bidding windows and prioritisation criteria, with outputs published openly where appropriate, contributing to the global evidence base.[1] In contrast, K4DD is a rapid-response programme with a £5 million budget, offering - via a helpdesk - rapid reviews of evidence, facilitated learning events, and tailored evidence products. Delivered through a consortium led by the Institute of Development Studies, K4DD strengthens internal capabilities, fosters collaboration, and promotes evidence-informed decision-making. Together, EF and K4DD form a key part of FCDO's demand-responsive evidence offer, complementing each other in scale, speed, and scope.

The combined value proposition of these programmes lies in their ability to deliver timely, responsive and policy-relevant evidence that meets the specific needs of decision-makers. They support strategic alignment, foster a culture of evidence use, and contribute to more effective and inclusive UK aid and diplomacy. Their design ensures that outputs are not only high-quality and timely but also accessible, equitable, and grounded in diverse perspectives. By engaging a wide range of research partners such as NGOs, think tanks, civic societies, consultancies and academic institutions - including those from underrepresented regions and groups - the programmes enhance contextual relevance and promote technical capacity development.

Oxford Policy Management (OPM)'s Value for Money (VfM) approach emphasises explicit evaluative reasoning - making transparent, evidence-based judgments about how well resources are used. It integrates insights from evaluation and economics, using stakeholder engagement and rubrics to define what "adequate', 'good', or 'excellent' VfM means in context. The framework goes beyond cost metrics, incorporating the 5Es - economy, efficiency, equity, effectiveness and cost-effectiveness - to assess complex interventions. It is designed to support adaptive management, learning, and accountability across diverse development programmes.

Programme-specific definitions for each of the 5Es were developed through a structured process, with clearly defined criteria and standards, providing a transparent basis for judging VfM. The framework was informed by 19 key informant interviews and three stakeholder workshops, which brought together FCDO programme managers, researchers, and evidence users. These engagements ensured that the criteria and standards reflect both strategic

1 A few research requests are also directly sourced, as they are demand-based.

priorities and operational realities, incorporating diverse perspectives and fostering collective ownership of the framework and future VfM assessments made using it.

The VfM framework for all 5E's is summarised below:

Economy: The VfM framework defines *Economy* as the efficient use of resources to enable high-quality, relevant research. It assesses how well management systems control costs while maintaining responsiveness and rigour. Key resource categories include administrative oversight, technical advisory, supplier costs, operational expenses, and intangible assets like trust and political will. Good stewardship is demonstrated through streamlined governance, responsive decision-making, proportionate commissioning, cost optimisation, and light-touch reporting. Performance on economy is judged using benchmarks, timelines, and stakeholder feedback, with criteria tailored to project complexity.

Efficiency: The VfM framework defines *Efficiency* as the ability of funded projects to deliver high-quality, relevant outputs in a timely and coherent manner. It assesses how well programme processes - from prioritisation to delivery - translate resources into usable evidence. Key dimensions include strategic prioritisation of research questions, clarity and accessibility of outputs, flexible engagement with users, streamlined commissioning, and responsive administration. Programmes should also draw on institutional knowledge and feedback to improve delivery. Performance is judged using timelines, stakeholder feedback, and output quality, with criteria tailored to project scale and complexity.

Effectiveness: The VfM framework defines *Effectiveness* as the extent to which research outputs are used to inform policy, programming, diplomacy, partnerships, and strategy. It assesses both the frequency and significance of impact across individual projects and the programme as a whole. Key impact types include instrumental (policy or practice change), conceptual (shaping understanding and debate), and process use which includes both capacity building (enhancing users' ability to engage with evidence), and connectivity (strengthening networks for future collaboration). Effectiveness varies by project and is influenced by context, user needs, and the robustness of findings. Good performance is demonstrated through tangible policy influence, meaningful shifts in thinking, improved user capability, and sustained engagement between researchers and decision-makers. The framework recognises that some impact may be cumulative and relational, and that not all projects will achieve direct impact due to external constraints. Programme-level assessments should consider how outputs collectively contribute to better policy and programming and a more evidence-informed ecosystem, with expectations tailored to the nature and delivery model of each project.

Equity: The VfM framework defines *Equity* as the inclusion of diverse voices and fair distribution of research benefits, especially for traditionally marginalised groups. It assesses whether research processes and outputs reflect varied perspectives and reach underserved users. Key dimensions include diversity in programme users, commissioned topics, research partnerships, and accessibility. Equity is demonstrated through inclusive commissioning, integration of intersectional and non-dominant knowledge, and outputs that are publicly available and user-friendly. Performance is judged using diversity metrics, content reviews, and stakeholder feedback. The framework encourages equity to be embedded across the programme lifecycle and assessed alongside other dimensions of value.

Cost effectiveness: The VfM framework defines *Cost-Effectiveness* as the ultimate impact and value of evidence generated by the programmes—on development priorities, UK diplomatic goals, and the culture of evidence use within FCDO and its partners, and through them on final improvements for the beneficiaries of these activities — and whether this justifies

the resources invested. Key dimensions include the value assigned by stakeholders and the value and the significance of impact compared with cost. Stakeholder value reflects how users perceive the relevance, merit, and distinctiveness of the programmes relative to alternatives. This includes satisfaction, continued engagement, and the uniqueness of the service provided. Impact significance is assessed through scope (reach), depth (scale of change), and equity (distribution of benefits). While some impacts may materialise beyond project timelines, the potential scale and relevance of change remain central to judging value. Performance is judged through stakeholder feedback, evidence of strategic influence, and the reach and depth of benefits. The framework encourages a balanced view that considers both perceived value and beneficiary impact.

To support implementation, the framework includes recommendations for data collection. Proposed data sources include a user survey, stakeholder interviews, programme dashboards, financial records, commissioning logs, QA reports, strategy documents, and accessibility analytics. The framework supports both formative[2] and summative[3] assessments, providing transparency and fostering learning. It promotes a balanced, portfolio-level approach to evaluation, acknowledging that programmes often face trade-offs between competing priorities such as depth versus breadth, innovation versus standardisation, thoroughness versus timeliness, and cost versus quality. By enabling ongoing improvement, the VfM framework offers a robust tool for enhancing the strategic and operational effectiveness of these two demand-driven evidence programmes.

[2] Formative assessments are evaluations of ongoing programmes, used to improve processes or outcomes during implementation.
[3] Summative assessments are final evaluations conducted after a process or project concludes. They often assess overall effectiveness or impact, as well as making judgements against other evaluation criteria.

Table of contents

List of Figures, Boxes and Tables

List of abbreviations

DFID	Department for International Development
5Es	Economy, Efficiency, Effectiveness, Equity and Cost-effectiveness
EF	Evidence Fund
FCDO	Foreign, Commonwealth and Development Office
FM	Fund Manager
HMG	His Majesty's Government
IDI	In-depth Interviews
IDS	Institute of Development Studies
K4DD	Knowledge for Development and Diplomacy
KII	Key Informant Interview
MEL	Monitoring, evaluation, and learning
MoU	Memorandum of Understanding
ODA	Official Development Assistance
OPM	Oxford Policy Management
PwC	PricewaterhouseCoopers
RED	Research and Evidence Directorate
RfP	Request for Proposals
ToR	Terms of Reference
ToC	Theory of Change
VfM	Value for Money

1 Introduction

This document defines a Value for Money (VfM) framework for two FCDO programmes that produce evidence for FCDO staff on demand: the Evidence Fund (EF) and Knowledge for Development and Diplomacy (K4DD). It also identifies data sources that could be used in applying the framework. The framework was developed by Oxford Policy Management (OPM) in consultation with key programme stakeholders. The objective of the work was to develop a clear and systematic VfM assessment approach for these two FCDO demand-responsive research programmes. It may also be relevant for other similar programmes, with adaptation.

Demand-responsive research services are essential for enabling informed decision-making and enhancing the effectiveness of FCDO's operations in various global contexts (Achillini, H., & Burge, R. 2024).[4] These services provide up-to-date and relevant information relatively quickly, which supports making informed decisions in fast-paced and dynamic environments. The services often involve collaboration with research institutions, think tanks, and other experts. They help decision-makers navigate complex global challenges by offering evidence that is relevant, practical, and aligned with the UK's development, diplomacy, and strategic goals. This kind of research informs policy, strengthens partnerships and supports the UK's role on the world stage by furthering its foreign policy priorities.

A comprehensive framework for Value for Money (VfM) assessments will address the FCDO's 5Es (Economy, Efficiency, Effectiveness, Equity, and Cost-effectiveness) and may extend to address a range of other VfM criteria if needed, including those with difficult-to-measure and difficult-to-value intended outcomes. It must be systematically grounded in a clear value proposition, ensuring a deep understanding of how value is created given the specific context. Incorporating diverse stakeholder perspectives is essential to giving voice to relevant parties and fostering inclusivity in decision-making. Additionally, since the programmes yield complex outcomes that are challenging to quantify, the assessment should integrate both qualitative and quantitative data.

OPM has worked with a core group of stakeholders and conducted key informant interviews with programme stakeholders to develop a VfM framework for demand driven research programmes. The plan entailed:

1. Document review of the business cases for EF and K4DD, annual reviews, the Monitoring, Evaluation, and Learning (MEL) framework for K4DD, and impact stories for EF (including an evaluation of EF support to the FCDO during the 2023 Nigerian elections[5]).
2. 19 Key Informant Interviews with 20 stakeholders – a summary of the key issues emerging from these is included as an annex in this report.
3. Three workshops with key programme stakeholders, including FCDO research and evidence advisors, programme managers, and evidence users (an annex with details of these workshops and discussion points is included in the report).

4 Achillini, H., & Burge, R. (2024, July 5). Demand-responsive research support to the Foreign, Commonwealth and Development Office during the 2023 Nigerian elections: an evaluation using outcome harvesting. Available at: https://doi.org/10.31235/osf.io/8wze7
5 Ibid.

Glossary of key terms

- **Research and evidence advisers:** Research and evidence advisers are FCDO staff, who lead the delivery of the Evidence Fund. Advisers provide end-to-end support for commissioning studies, working to identify evidence needs, evaluating bids or requests from potential users against the FCDO's priorities, managing procurement and onboarding of research partners, supporting quality assurance of evidence products, and driving research uptake.
- **Research and Innovation Hubs:** Research and evidence advisers are embedded within regional teams, referred to as Hubs, and maintain close ties with FCDO country offices. They are as follows: South and Southeast Asia Hub, West Africa Hub, East Africa Hub, Southern Africa Hub, and the UK Hub.
- **Programme managers:** The term programme manager is used to refer to the FCDO teams responsible for the delivery of the Evidence Fund and K4DD programmes, namely their Principal Responsible Owner(s) and Senior Responsible Owner(s). In the case of K4DD, this also includes the IDS personnel who manage delivery on behalf of the consortium.
- **Fund Manager:** With reference to the Evidence Fund, the Fund Manager is a contracted organisation, which manages the programme's procurement, payment-related processes, and financial reporting on behalf of the FCDO.
- **Research partners or suppliers:** All parties contracted to deliver research and evidence outputs for the FCDO are referred to as research partners or suppliers. These may include, in the case of Evidence Fund, research and academic institutions, universities, for-profit consulting and research agencies, NGOs, civil society organisations, think tanks, policy and research networks, and semi-government bodies. In the case of K4DD, this refers to the delivery consortium made up of Institute of Development Studies (IDS), the University of Birmingham, Liverpool School of Tropical Medicine, the University of Manchester, the Royal United Services Institute, and the Association of Commonwealth Universities.
- **Evidence users or research customers:** This term refers to FCDO staff or stakeholders who engage with, commission, or apply the research and evidence produced under the Evidence Fund or K4DD to inform their work. They may include FCDO country advisers, policy teams, programme managers, and senior leadership across central departments or country posts.

Draft elements of the VfM framework, focused on the value proposition and possible assessment criteria, were developed iteratively, based on information from the document review, the first workshop and 19 KIIs. This was shared with the core group for comments and discussed in the second and third workshops. This document provides the final framework with recommendations on possible data sources. It also presents OPM's approach to VfM and gives a brief description of the programmes to provide context.

The framework was designed to inform VfM assessments at the programme level, that is, to assess each programme as a whole. It was designed to be broad and flexible enough to be used with the EF or K4DD, which means that some individual elements may be less applicable to a particular programme. For example, the EF employs a strong prioritisation process to ensure that its more expensive outputs are produced only for the issues of most strategic importance to the organisation. K4DD offers its much lower-cost, services to all comers on a first-come, first-served basis. The two programmes will therefore be rated differently on sub-criteria for effectiveness and equity – with EF rated better on 'strategic relevance and structured prioritisation' (part of effectiveness) but lower on 'catering for a diverse user base'

(part of equity). By not limiting the framework to sub-criteria on which both programmes place similar priority, it provides a basis for a high-level comparison between the two across all five criteria. The assessment of cost-effectiveness should help to bring out the differences between the two programmes in terms of their overall complementarity, impact and value to users.

The framework was not designed to be used at the project or study level, but aspects of the framework could be applied to make assessments at either the individual project level or the portfolio level, to make comparative assessments between different demand-driven evidence programmes. In general, effectiveness criteria are most relevant to project-level assessments and cost-effectiveness criteria are most relevant to portfolio level comparisons; specific applications are noted in the text below. Economy and efficiency criteria are mostly suitable to the programme level. The extent to which individual projects in a programme meet expectations for economy, efficiency and equity should be managed by programme level policies and processes. For example, optimised costs (an economy sub-criterion) can be controlled at the project level through well-designed programme level policies on eligible costs. Likewise, clear programme level guidance and quality assurances processes can ensure that outputs from individual projects are accessible and coherent to non-technical audiences.

The assessment of effectiveness will vary between individual projects in a programme, as outcomes (particularly in terms of instrumental and conceptual impacts) will vary between projects. In these circumstances, a programme level assessment will need to include an aggregation or representative sample of project-level assessments.

Cost-effectiveness criteria could potentially be used to inform comparisons between programmes. Some of the sub-criteria under cost-effectiveness involve estimations of the long-term impact of evidence-informed decisions, which is ultimately required to determine if the cost of inputs have produced a return.

1.1 OPM's approach to VfM

OPM's approach to assessing VfM provides a robust and transparent framework for evaluating how well resources are used and whether the value created justifies the investment. This interdisciplinary approach combines insights from evaluation and economics, emphasising explicit evaluative reasoning through the use of transparent criteria and standards. By integrating both quantitative and qualitative evidence and encouraging participatory engagement with stakeholders, the approach supports evidence-based judgements that are contextually relevant and aligned with broader monitoring, evaluation, and learning (MEL) processes.[6]

Within this framework, VfM criteria define the key dimensions of program performance that underpin good resource use. At a broad level, they specify the aspects of performance that require evidence to support an evaluative judgment of VfM.

FCDO typically assesses the VfM of its investments against the '5Es'- Economy, Efficiency, Effectiveness, Equity, and Cost-effectiveness. FCDO's definitions of these criteria are given in **Table 1** and their relationship with a generic project results chain is shown in **Figure 1**.

6 King, J., Wate, D., Namukasa, E., Hurrell, A., Hansford, F., Ward, P., Faramarzifar, S. (2023). *Assessing Value for Money: the Oxford Policy Management Approach*. Second Edition. Oxford Policy Management Ltd.

Table 1: Generic definitions of the 5Es (DFID, 2020)[7]

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 are agents invest in it?

OPM generally uses a somewhat broader understanding of these criteria, allowing for a more tailored analysis of a programme's VfM. The framework development process begins with 'value proposition' questions asked to key stakeholders. These are designed to prompt consideration about what it would look like for the programme to be a good steward of resources (economy), deliver its outputs appropriately (efficiency), achieve outcomes (effectiveness), create enough value to justify the resources invested (cost effectiveness) and do so equitably (equity).[8] The value proposition questions cover all the 5'Es'. The value propositions inform the development of the VfM framework.

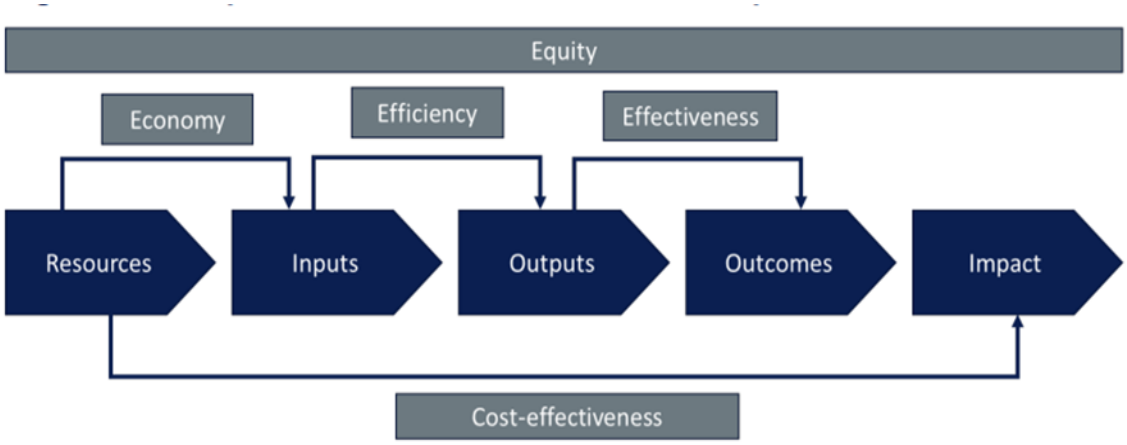

Figure 1: 5Es related to the results chain

7 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: Smart-Guide - Approach-to-Value-for-Money_External.pdf
8 King, J. (n.d.) *'Value Propositions – Part 2: Clearing the Path'*, *Juliankingnz Substack* [online] Available at: Value propositions (part 2) - by Julian King

As with any evaluation, VfM criteria must be contextually determined, reflecting the specific attributes of a project, programme, or policy that contribute to optimal resource allocation and impact. A comprehensive VfM assessment should address the following core questions:

1. How economically and efficiently have resources been utilised?
2. What value has been generated through this investment?
3. Does the value created justify the resources expended?
4. How can resource use be optimised to enhance impact?

Beyond making summative judgements, VfM assessments should identify opportunities for improvement - whether by refining an existing intervention or considering an alternative approach. The goal is to maximise the effectiveness of resource use in achieving intended outputs, outcomes, and value. Therefore, VfM criteria should include aspects of performance that facilitate meaningful learning and inform decision-making, ensuring that evaluations remain utilisation-focused (Patton and Campbell-Patton, 2021).[9]

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).[10] We have developed a set of standards to provide generic definitions of different levels of performance, which are detailed in our Guide (p. 25).[11] The programme-specific standards we developed for this framework are aligned with our generic standards, providing consistency across VfM frameworks in the underlying meaning of terms like 'excellent' and 'good'.

The practical approach for designing, undertaking, and reporting a VfM evaluation follows 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. The key steps involved in explicit evaluative reasoning are summarised in **Figure 2** 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. This report describes the VfM framework, it does not provide a VfM assessment using that framework. The report provides some examples of how the rubrics may be applied. These are illustrative and should be more fully developed and refined when the VfM assessment is undertaken, in the light of further document review, the data gathered and analysed, and further engagement with key stakeholders.

9 Patton, M. Q., & Campbell-Patton, C. E. (2021). *Utilization-focused evaluation* (5th ed.). SAGE Publications.
10 Davidson, E.J. (2014) 'Evaluative reasoning', Methodological Briefs: Impact Evaluation 4, UNICEF Office of Research, Florence [online]. Available at: https://www.unicefirc.org/publications/749-evaluative-reasoning-methodological-briefs-impact-evaluationno-4.html
11 King, J., Wate, D., Namukasa, E., Hurrell, A., Hansford, F., Ward, P., Faramarzifar, S. (2023). *Assessing Value for Money: the Oxford Policy Management Approach*. Second Edition. Oxford Policy Management Ltd.

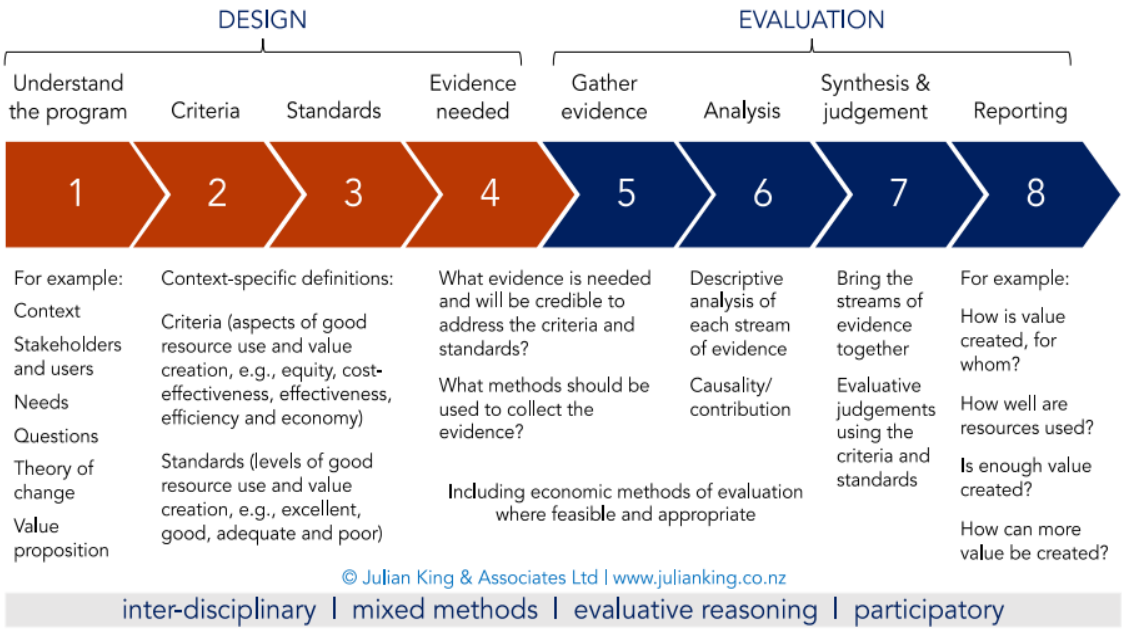

Figure 2: The eight step VfM approach

1.2 Overview of the programmes

Evidence Fund

The Evidence Fund commissions high-quality, demand-driven primary and secondary research, evaluations, and evidence syntheses. It aims to ensure that FCDO staff have access to timely, context-specific, and operationally relevant evidence which meets their specific needs, which may range from designing and adapting development programmes, to informing policy positions, responding to emerging global challenges, and strengthening strategic relationships with partner countries. Evidence generated through the Evidence Fund plays an important role in shaping the UK's strategic portfolio, including country-level and regional business plans, and supporting diplomatic engagements with partner governments.

In addition to its official development assistance (ODA) remit, the Evidence Fund also supports non-ODA research (following an addendum to the programme's business case in January 2024), which advances the UK's humanitarian, security and foreign policy objectives. This includes informing Science & Technology priorities and contributing to the UK's growth ambitions.

Launched with a budget of up to £58 million over a seven-year period, from 2020 – 2027 the Evidence Fund consolidated several (previously separate) demand-responsive research programmes under one streamlined platform to improve efficiency, coherence, and learning across the FCDO's evidence functions. The consolidated programme seeks to minimise the duplication of effort in contract management and procurement processes, while enabling faster and more strategic commissioning of research, synthesis, and evaluation work. The intended impacts of the Evidence Fund are as below.

- More effective, evidence-informed FCDO policy and programming, leading to better global development outcomes and the furthering of UK's growth and foreign policy

objectives and improved value for money for FCDO's research and evidence commissioning
- Increased use of high-quality evidence in UK diplomacy and strategic partnerships
- An expanded global knowledge base, with public research outputs available to practitioners, partner governments, and multilateral organisations.

The programme's intended impact is underpinned by three key outcomes:

- A stronger understanding and use of evidence across FCDO and relevant external audiences,
- Enhanced policy and decision-making within FCDO and external organisations, based on credible, timely and context-relevant research, and
- Improved cross-organisational learning

Delivery of the Evidence Fund is led by a dedicated group of research and evidence advisers, together with a Senior Responsible Owner (SRO) and Programme Responsible Owner (PRO) within the FCDO's International Science and Technology Directorate. The research projects are managed by regional Research and Innovation hubs in Asia and Africa along with UK-based teams, namely the UK Hub, Evidence Syntheses unit, Evaluation Unit. They work together to identify evidence needs among potential users, support quality assurance of evidence products and drive research uptake. FCDO staff are supported by a Fund Manager (FM), which manages procurement, payment-related processes, and financial reporting.[12] The programme serves research requests from HMG's overseas missions across Asia, Africa, Europe, Latin America and MENA and its strategy and policy teams.

While it is not designed as a capacity-strengthening initiative, the Evidence Fund contributes to strengthening research ecosystems in Asia and Africa. In line with their localisation agenda, FCDO hubs prioritise partnerships with local research suppliers and engage them through targeted commissioning, advertising upcoming research projects on social media platforms (LinkedIn and X) and leveraging professional networks. These efforts support technical capacity development for local research partners while helping FCDO obtain more contextually relevant research outputs.

The Fund's demand-led process allows FCDO staff across countries, regions, and central departments to submit evidence requests. These requests are assessed against a range of criteria by the FCDO research advisers, who prioritise evidence needs that are cross-cutting, underserved, aligned with HMG priorities and partner countries or considered high priority for furthering FCDO's growth and development objectives. Periodic "bidding windows" are launched to proactively solicit research requests from potential users. The key selection criteria for determining funding allocation include:

- Strategic alignment with FCDO and HMG priorities
- Connectedness of the research with previous, ongoing, or future activities
- Quality of the research question(s) posed by the user
- Clarity of plans to apply the findings from the requested research in decision-making
- Usefulness and degree of unmet demand and/or evidence gaps
- Feasibility of the study requested
- Time-sensitivity and responsiveness to real-time decision-making needs
- Any anticipated risks (political, reputational, diplomatic, etc.) applicable for Official Sensitive projects.

12 Currently contracted to PricewaterhouseCoopers (PwC)

Once a user request has been allocated funding and chosen to be taken forward, the Fund Manager and research advisers manage the contracting of a research supplier through a competitive tendering process. Suppliers for the Evidence Fund include research and academic institutions, universities, for-profit consulting and research agencies, NGOs, civil society organisations, think tanks, policy and research networks, and semi-government bodies.

Outputs including reports, multilingual policy briefs, presentations, infographics are published on the gov.uk website or Evidence Fund library, contributing to the global evidence base, with some exceptions for 'Official Sensitive' studies. The programme also aims to publish all ODA-funded outputs in open, editable formats (e.g., .odt) to support broader public access (including users of adaptive technologies like screen readers).

Knowledge for Development and Diplomacy (K4DD)

Knowledge for Development and Diplomacy (K4DD) is a rapid, demand-led evidence and learning service funded by the FCDO. The programme aims to enhance the use of high-quality, relevant and timely evidence in FCDO's development and diplomacy policy and programmes, and create opportunities to strengthen knowledge-sharing, learning and networks among FCDO staff and K4DD partners. Launched in October 2023 and running through to March 2027, K4DD is managed by the FCDO's International Science and Technology Directorate (Global Science Department). The programme is designed to improve FCDO staff capabilities to use evidence more effectively, foster collaboration across policy areas, and stimulate internal and external learning. It aims to address pressing developmental and diplomatic challenges through the synthesis, dissemination and application of existing knowledge. The expected outcomes of K4DD include:

- A learning-oriented organisational structure within FCDO
- Sound, evidence-informed decision-making and internal consensus on directions, through increased uptake of evidence
- Strengthened evidence networks and alliances for decision-making, through enhanced collaboration between FCDO and academic, civil society, and policy partners
- Improved development and diplomacy outcomes through more effective policy design and implementation

K4DD aims to supplement and build on the FCDO's in-house expertise, providing continuing professional development support to FCDO staff by bringing academic expertise and perspective to challenge, refresh, and widen their knowledge.

K4DD is delivered through an Accountable Grant worth £5 million and implemented by a consortium of leading development and diplomacy knowledge institutions, with the Institute of Development Studies (IDS) serving as the lead delivery partner. The consortium also includes the University of Birmingham, Liverpool School of Tropical Medicine, the University of Manchester, the Royal United Services Institute, and the Association of Commonwealth Universities. The consortium operates a standing team of researchers, with access to a wider pool of global experts in different thematic and geographic areas. The K4DD delivery partners are responsible for all aspects of implementation, including demand intake, research production, expert engagement, communications, and performance monitoring. The FCDO team, which includes a dedicated Programme Responsible Officer (PRO), provides oversight to ensure compliance and quality delivery.

Designed as a free-at-point-of-use service for all FCDO staff, the K4DD offer includes three components:

1. **Helpdesk Service:** Allows staff to request rapid evidence products such as:
    - Rapid Evidence Reviews, i.e. written syntheses of evidence on a specific topic or issue, which are produced within four weeks
    - Emerging Issues Reports, which go into greater breadth and depth and are completed within eight weeks
    - Rapid Bibliographies i.e. a list of sources with short abstracts
2. **Facilitated Learning Events** including Expert Challenge Sessions (one-off sessions) and Evidence and Policy Clinics (which take place over several months) to encourages peer learning and evidence-based dialogue on policy and programming related questions. The programme aims to conduct 3-5 Challenge sessions and 2-3 Evidence and Policy Clinics every year.
3. **Learning Products** i.e. outputs (such as presentations or resource packs) created to communicate selected outputs from the above-mentioned services to FCDO staff and externally. Outputs are also made public (where suitable), contributing to the global development knowledge base.

Requests for support through K4DD are reviewed against a defined set of criteria which includes clarity of the request(s), the feasibility of addressing the request within the scope of service offered by K4DD, and available technical expertise to deliver against users' expectations. Requests which are deemed beyond the scope of K4DD are redirected to other appropriate evidence services (including the Evidence Fund) or other internal mechanisms. K4DD places a strong emphasis on quality assurance, continuous learning, and adaptive management. The programme also aims to publish all non-sensitive outputs and learning materials through its website, further strengthening transparency, accessibility, and the reach of its evidence base.

The two programmes complement each other: while K4DD generally provides users with rapid evidence using a small budget per output (typically, 5.5-6 person-days), research commissioned through the Evidence Fund is more expansive in scope and budget per output (typically 3-6 month-long projects led by research teams).

2 Value proposition: to whom and in what ways are the programmes important?

It is useful to consider both separate value propositions and one statement that aims to cover both programmes. Proposed statements are as follows.

2.1 Evidence Fund: Overall Value Proposition

The Evidence Fund provides decision-makers with timely, high-quality, and thematic and policy-relevant evidence that directly responds to their specific needs and priorities. Operating as a demand-led programme ensures that it responds to immediate, real-world questions faced by UK government teams and partner countries. It supports the commissioning of research, evaluation and synthesis to inform and improve HMG strategy, programming, policy, diplomacy, partnerships and other internal decisions, and strengthens the capacity for and culture of evidence use. Production by a consortium of expert providers and management and guidance by FCDO research and evidence advisors and programme managers, ensures that outputs are of high quality, accessible, innovative where appropriate, produced efficiently and meet the needs of a range of users, as well as involving a range of evidence producers (i.e. research partners), encompassing diverse perspectives and addressing equity issues whenever appropriate. The outputs help ensure UK aid and diplomacy are more relevant, effective, and impactful and that FCDO and its partners, including country governments and multilateral bodies, have a growing evidence base and are better able to use evidence from a range of sources over the medium term. In addition, most outputs are shared in the public domain and thereby improve the global knowledge base and inform decisions by a range of other actors.

2.2 K4DD Programme: Overall Value Proposition

The Knowledge for Development and Diplomacy (K4DD) programme offers FCDO and its partners demand-led, rapid access to high-quality evidence and learning. It brings together and shares existing evidence - through low-cost rapid evidence reviews, learning events (i.e. expert challenge sessions and evidence and policy clinics), a helpdesk and learning products – via a consortium of expert academic institutions. Production by this consortium and management and guidance by FCDO research and evidence advisors and programme managers ensures that outputs are of high quality, accessible, produced rapidly and efficiently and meet the needs of a range of users, as well as involving a range of evidence producers, encompassing diverse perspectives and addressing equity issues whenever appropriate. This work directly informs and improves strategy, programming, policy and other decisions, as well as strengthening the evidence base and capacity for and culture of evidence use and so helps ensure UK aid and diplomacy are more relevant, effective, and impactful and that FCDO and its partners are better able to use evidence from a range of sources over the medium term. In addition, outputs or learning products shared in the public domain improve the global knowledge base and inform decisions by a range of other actors.

2.3 Combined value proposition

The Evidence Fund and the Knowledge for Development and Diplomacy (K4DD) programme together provide FCDO and its partners with timely, high-quality, thematic and policy-relevant

evidence and learning that directly responds to their specific needs and priorities. Operating as demand-led programmes ensures that research and evaluation activities respond to immediate, real-world questions. The programmes are complementary, with K4DD providing low-cost, rapid access to existing knowledge, while EF generates new research, evaluations and synthesis on a longer timescale. Production by a consortium of expert providers and management and guidance by FCDO research and evidence advisors and programme managers, ensures that outputs are of high quality, accessible, innovative where appropriate, produced efficiently and meet the needs of a range of users, as well as involving a range of evidence producers, encompassing diverse perspectives and addressing equity issues whenever appropriate. This evidence informs and improves HMG strategy, programming, policy, diplomacy, partnerships and other decisions, as well as helping to strengthen the evidence base and capacity for and culture of evidence use, helping ensure UK aid and diplomacy are more relevant, effective, and impactful. In addition, outputs inform UK partners (such as country governments and multilateral bodies), and most outputs are shared in the public domain and thereby improve the global knowledge base and inform decisions by a range of other actors.

2.4 VfM Criteria and Standards

The VfM framework is based on a set of criteria, against which performance standards are defined and VfM judgements made. These criteria should identify the key elements of programme value and how it is delivered. The criteria and standards are described in the following sections, according to the 5Es. **Against each criterion, we have also provided examples on how the programme may be judged on it. These examples are indicative and can be tailored as per the programme SOP. Error!** Unknown switch argument. provides a brief overview of how these criteria have been interpreted in the context of the EF and K4DD.

Table 2: Specific definitions of the 5Es for demand driven research programmes

Criterion	Definition
Economy	How well do the management systems and processes control costs while enabling high-quality, relevant research projects?
Efficiency	How well do funded projects produce high-quality, relevant outputs?
Effectiveness	How well are the outputs used to inform policy, programming, diplomacy, partnerships and strategy?
Equity	Do the processes for prioritising, commissioning, conducting and using research involve a sufficient diversity of experiences, voices, and perspectives, including those that are traditionally marginalised? Are the benefits of the research fairly distributed, and do they meaningfully reach and represent underserved or disadvantaged groups?
Cost-effectiveness	What is the ultimate final impact and value of the evidence generated by the programmes on development priorities, the UK's diplomatic and strategic goals and the culture of evidence use within FCDO and its partners, and does it justify the inputs that are invested in it?

The framework has been designed recognising that all programmes face tensions and trade-offs between competing priorities. In some cases, these trade-offs are recognised specifically in the VfM framework. In other cases, when applying the VfM assessment users will need to actively consider the balance struck between priorities, both within and between criteria, in

making an overall assessment of VfM. We expand on these trade-offs in the context of K4DD and EF in Section 8.5.

3 Aspects of value: Economy

Economy is defined as: how well do the management systems and processes control costs while enabling high-quality, relevant research projects?

3.1 Potential criteria for economy

The development of the criteria and standards for economy is informed by an understanding of which resources are invested in the programme and what good stewardship of those resources looks like.

The main types of resources invested in the programme are summarised in **Error! Reference source not found.**Table 3. Please note that these are not aligned with the criteria which have been developed for Economy.

Table 3: Types of resources invested in FCDO demand-driven evidence programmes

Cost / investment category	Details
Administrative, Finance & Programme Oversight (FCDO and Fund Manager)	Administrative budgets for project commissioning (procurement), and overall operational and financial oversight and reporting support. Covers programme-related governance and operational coordination, supported by PwC / IDS for management, contracting and compliance.
FCDO Technical Advisory & Programme Delivery	Knowledge and expertise of technical advisors in research commissioning, TOR design, supplier selection, study oversight, research and technical advisory into reports, dissemination, and internal reporting.
Research Supplier Costs	Fees and expenses[13] in terms of time and expertise of researchers and institutions in delivering research/evaluation outputs, managing consortia, shaping methodologies, and participating in co-creation and review processes.
Other Operational Expenses	Travel, lodging, and logistical costs tied to in-person collaboration and programme delivery.
Intangibles: Network Capital & Relationships including political will	The often-unseen value of long-term trust, credibility, embedded relationships, and institutional memory among commissioners, suppliers, and partners and political will that is essential for effective, responsive programming.[14]

Good stewardship of resources can be described in terms of the application of the following principles. Box 1 demonstrates the difference in EF and K4DD on the criteria.

The often-unseen value of intangibles such as long-term trust, credibility, embedded relationships, and institutional memory among commissioners, suppliers, and partners along

13 Fees cover expertise costs of researchers and expenses include all data validation workshops, dissemination events, field visits, etc.
14 To reflect the unseen and uncosted value of trust and institutional memory in VfM, we can include qualitative narratives and case examples that show their impact on programme responsiveness. Stakeholder perception surveys can further quantify credibility and embedded relationships, linking them to tangible outcomes.

with political will - essential for effective, responsive programming - is also considered as good stewardship of resources.

- **Administrative and governance structures are streamlined**, with clear roles and minimal duplication. Demand identification and project shaping are timely, evidence-informed, and strategically aligned. Management systems enable agile implementation and oversight with minimal waste[15] and transaction costs. Some examples on which these can be judged are as follows: For administrative & governance structures, 'excellent' could mean FCDO rolls out the call for evidence requests and reaches a wide range of FCDO regional and country teams, implements standardised assessment of all proposals against a clear set of criteria, and selects projects for commissioning and announces results in a timely manner. 'Good' could mean that the process meets these expectations in most respects, but fails on a single element e.g. in reaching only a limited set of FCDO teams and so missing some potential users. 'Adequate' when the process falls short on a number of elements, while it would be 'Poor' if it falls short on all of them.
- **Timely and streamlined commissioning process:** The commissioning process is timely and proportionate to the scale and complexity of the project. It is guided by clear procedures that streamline project requests and assessments. Structured commissioning and delivery frameworks ensure that project requests are assessed consistently and efficiently - minimising delays and reducing administrative burden. This could be judged, for example by the number of small-scale rapid studies which are commissioned using a light-touch process, particularly for K4DD, while larger projects follow a more rigorous path particularly for EF. Illustratively, this could be judged as poor if it takes more than a month from supplier selection to completion of due diligence and contracting, adequate if it takes between 3-4 weeks, good if it takes around 2 weeks and excellent if it takes less than 2 weeks. Some of the criteria can also be assessed in user surveys or KIIs with relevant stakeholders on how easy the systems are to interact with.
- **Optimising the cost of research outputs:** Costs are actively optimised by matching the right level of effort to each task, balancing speed with rigour, and utilising suppliers' databases that have been built by regional teams and posts. It is to be noted that there could be trade-offs between the best knowledge/skills of the research supplier and their availability within the project timeline, as well as other considerations affecting this. An example to judge these is benchmarking i.e. comparison of cost per evidence product or consultation against similar projects or historical data. This is particularly important as K4DD would have very different unit costs for rapid reviews, as compared to detailed studies undertaken by EF making standardisation challenging. The performance standards would therefore be based on thresholds which are relevant to the type of output, taking cognizance of speed and rigour. For instance, a rapid review would have a shorter timeline/budget than a systematic review or an evaluation. While a direct comparison of the cost per evidence product may be challenging due to differences in scope, methodology, and timelines, benchmarking can still offer some broad insights, when combined with other sources of information. By grouping similar types of products (e.g., rapid reviews vs. systematic evaluations), one can establish indicative cost ranges and performance thresholds that reflect the nature and rigour of each output. In this context, the term *programme* refers to the overall initiatives such as EF and K4DD whose performance is being evaluated. The assessment focuses on whether the quality of outputs is sufficient and whether the programme's total cost aligns with the average cost of producing similar outputs across comparable programmes. This involves reviewing the average costs of outputs by type (e.g., reports, datasets, tools), and calculating the

15 Waste refers to the inefficient use of resource such as time, money, personnel, or systems that do not directly support strategic objectives. It includes duplication, unnecessary processes, underutilized assets, and costs due to poor planning/organisation.

deviation from those benchmarks. A programme can be judged as excellent if the quality of outputs is sufficient and it is within 5% of the average budget for similar outputs, good if it is no more than 5-10% above the average budget, adequate if it is less than 10-20% above the average budget and poor if greater than 20% more than the average budget of similar outputs, and if there is no justification for these deviations.[16]

- **Transparent fund management and streamlined reporting:** Ensuring transparency, light-touch and streamlined reporting allows researchers/research teams to focus on delivering quality outputs rather than excessive project management documentation, without compromising accountability. This can be evaluated based on whether reporting templates are simple and demands made are proportionate to project size. Here, surveys or interviews with research partners can be used to assess whether researchers felt reporting requirements were reasonable and allowed them to focus on delivery always (excellent), at most times (good), sometimes (adequate) and never (poor). Research partners may also report on whether they believe FCDO project management meetings were a time-efficient means to manage and report on the project. Another illustration could be *excellent* would be if Fund Managers and FCDO Programme Managers consistently maintain up-to-date fund management spreadsheets tracking costs against budgets and hold monthly oversight meetings as standard practice. A *good* example would be if these spreadsheets and meetings are generally maintained, with only occasional lapses. An *adequate* example would be if spreadsheets exist and meetings are scheduled, but without consistent follow-through or updates.
- **Complementarity:** Past investments are treated as reusable assets, with teams drawing on existing insights and partnerships to avoid duplication and strengthen complementarity across FCDO platforms, making sure that requests go to the most appropriate evidence source. This can be judged by reviewing whether new investments filled gaps rather than overlapping with existing ones. A few indicators that can support are percentage of new projects that reference or build on past investments and number of duplicated efforts identified and avoided. It has emerged in discussions that around 15 - 20% of new research projects build on past research work commissioned by EF, most well aligned to the local country needs or FCDO's strategic needs. This can be considered as narrative evidence showcasing one element of complementarity adjudged as excellent. Another illustration for excellent could be if technical advisers regularly work with known research customers / research partners to enhance the speed and quality of commissioned work.

3.2 Justification

The refined criteria for economy in the VfM framework are grounded in ensuring optimal resource allocation and stewardship through structures and processes that promote economic use of resources while maintaining the integrity and efficiency of research outputs. The KII insights highlight the tangible financial commitments - such as researcher time, fund management, and administrative overhead - that shape programme expenditures. Meanwhile, the workshop discussions emphasised the importance of balancing administrative and research costs to focus resources on final outputs, while maintaining structured operational frameworks for efficiency. Outsourcing programme management functions to IDS (K4DD) and PwC (EF) helps in managing operations, allowing FCDO to commission and deliver research outputs, improving delivery timelines and minimising FCDO's administrative burden. Additionally, a light-touch approach to reporting ensures researcher partners remain focused

[16] These assessments must take account of context, user requirements and variations in costs faced. If average output costs are found to be higher than benchmarks, a qualitative assessment as to whether these higher costs are justified by such factors should be undertaken, for example through a review of budgets and outputs of a sample of research projects.

on substantive work rather than excessive documentation, consistent with an agile and effective fund management system.

Prioritising user engagement and network capital underscores the significance of trust-based relationships, ensuring long-term partnerships and effective stakeholder collaboration. Both the KIIs and workshop insights highlight the necessity of refining resource distribution to align with evolving programme objectives.

3.3 Performance standards for economy

How well do the management systems and processes within demand driven programmes control costs while enabling high-quality, relevant research projects?

Performance	Adequate	Good	Excellent
	The programme has **administrative and programme management processes** in place, with some shortcomings in project handling.	The programme has clear **administrative and governance processes** in place. There are systems for demand identification, project shaping, and management with minimal waste, with minor exceptions.	The programme has **lean and clear administrative and governance structures**. There are systems for demand identification, project shaping, and management, with minimal waste
	There are **processes in place for streamlined commissioning**, inputs are mostly proportionate to project needs, and effort balances rigour with responsiveness, with scope for improvement.	**Commissioning is mostly streamlined**, inputs are proportionate to project needs, and effort balances rigour with responsiveness with minor exceptions	**Commissioning is streamlined**, inputs are proportionate to project needs, and effort balances rigour with responsiveness.
Sub-criteria	**Operational costs are generally necessary**, but with some unplanned and or disproportionate costs, costs are suboptimal with scope for improvement.	**Operational costs are generally** necessary, planned and proportionate, costs are aligned with level of effort, with minor exceptions.	**Operational costs** are necessary, planned and proportionate, costs are aligned with the level of effort.
	The programme has **processes in place for transparent fund maintenance and oversight** – but with room for improvement.	The programme has **transparent fund maintenance and oversight** - strong financial controls in place, but with some room for improvement.	The programme has **transparent fund oversight,** minimising bureaucratic overhead.
	Complementarity: The programme has only occasional overlap in leveraging existing networks – there is some duplication of efforts and reporting	**Complementarity**: The programme leverages on existing networks and ensures previous research generates ongoing value without excessive duplication, with minor exceptions	**Complementarity**: The programme leverages long-term professional collaboration and ensures previous research generates ongoing value without duplication
Performance will be rated as poor if any of the conditions of adequate are not met			

4 Aspects of value: Efficiency

Efficiency is defined as: how well do funded projects produce high-quality, relevant outputs?

4.1 Potential criteria for efficiency

Identification of priority areas and getting the right research questions

- **Strategic relevance, structured prioritisation and getting the right research questions:** Use appropriately structured bidding or decision processes to assess relevance, feasibility, and alignment with UK policy priorities, so resources are allocated to the most important work. Systems seek to identify and eliminate duplication and refer users to other services, where the scope or requested methodology is unsuitable for a particular programme. While structured prioritisation may be less applicable to K4DD, which operates on a first-come, first-served basis, EF applies a more rigorous prioritisation framework - evaluating proposals against strategic criteria to ensure coherence and policy alignment. However, K4DD also ensures the requested research questions are valid and conceptually robust, to prevent resources from being wasted. It should be noted that sometimes there may be an inherent trade-off between academic rigour and policy relevance. An example of how this can be judged: the programme demonstrates excellent efficiency, by using a proper systematic approach to quickly assess requests for strategic fit and duplication, redirecting unsuitable requests within a stipulated time. A good level is shown when the programme undertakes the review with clear criteria and some redirection of requests, ensuring reasonably timely and effective resource use. At an adequate level, the programme accepts proposals on a rolling basis but lacks a structured process, resulting in slower prioritisation and occasional inefficiencies.

Research commissioning and implementation

- **Accessibility and Coherence:** The extent to which research findings are accessible and presented with clarity and coherence, such that their logic and implications are understandable to non-technical audiences. For instance, excellent findings use plain language and are easily understood by general audiences. Good findings are generally clear but occasionally use potentially confusing technical terms. Adequate findings are those with good rigour or insights but presented in specialist language that hinders broader understanding. More examples of accessibility include systematic use of policy briefs for policymakers who may not have time to go through full reports and translated slide decks and policy briefs for uptake and absorption by local audience.
- **Quality outputs:** Research findings are delivered within a timeframe that enables their meaningful integration into decision-making processes (timeliness); are of sufficiently high quality and/or reviewed by external peer reviewers, as reflected in methodological rigour and analytical depth and, where appropriate, contain innovation and novel ideas; and demonstrate integrity by drawing on up-to-date, credible evidence and ethical research standards. Ethical compliance safeguards the integrity of findings, protects participants, and promotes transparency which are key attributes of high-quality research.[17] Ethical research avoids waste, duplication, and harm- ensuring that

[17] https://esrc.ukri.org/research/our-research-ethics/

resources are used responsibly. It also supports long-term impact by maintaining public confidence and stakeholder engagement. For country-level research studies, particularly those involving fieldwork or human participants, Institutional Review Board (IRB) approval is a critical requirement. In short-term or low-budget projects, trade-offs are often faced between doing something quickly and doing it more thoroughly. A high-quality output might mean it is useful, clear, relevant and reliable- even if it is not exhaustive. Rigour means using methods that are appropriate for the task, not necessarily the most complex. Programmes must balance speed, cost, and depth to get the best possible result within the constraints, while ensuring that the output if of a sufficient standard for the purpose it is being used. An example of judging one element of the quality of outputs can be excellent if timelines are always adhered to, good if they are adhered to at most times, adequate if sometimes and poor if never.

- **Collaborative engagement:** Work is based on partnerships where research advisers, users, and research partners work together to ensure that the output responds to users' needs in terms of content, format, and timing. This does not always require deep co-creation. Some users may prefer direct access to tailored evidence without extensive engagement, especially when their needs are well-defined. The commissioning process should therefore be flexible enough to accommodate a range of approaches: intensive collaboration when needed and streamlined delivery when appropriate. Here, excellence can be demonstrated by sufficient engagement to provide a thorough understand of users needs and a product that responds precisely to users' specific needs and has high user satisfaction; good performance implies meaningful consultation so that the output broadly reflects users needs ; adequate may be when alignment of the product with user needs is based largely on prior knowledge with insufficient interaction with the users to fully respond to user requirement though it responds to their broad area of interest. .
- **Optimised research administration, operational efficiency and technology integration:** Review and refine fund management processes to ensure and improve efficiency in procurement, administrative management and support to research partners with invoicing and timely payments, making use of technological improvements and AI (for plagiarism checks) where possible. This can be judged as excellent if the payments are always timely i.e. within 30 days, good if mostly within 30 days,adequate if they are only sometimes made within 30 days and poor if they are never made within 30 days or are sometimes delayed over 60 days.

Enabling factors

- **Stakeholder and institutional knowledge and continuous learning and improvement:** Foster effective, collaborative partnerships among research suppliers, research advisers and research users (FCDO stakeholders and external partners, if applicable) through proactive engagement, knowledge-sharing, and collaborative problem-solving. Leverage institutional knowledge to enhance decision-making, improve efficiency, and drive impactful outcomes. This can be done by using the organisation's experience to make better, faster decisions, thereby getting better results. Structured programme performance assessments including quarterly and annual reviews for both EF and K4DD, as well as outputs from the MEL components in EF and K4DD drive adaptive learning by actively seeking feedback, embracing new knowledge, and adapting to evolving challenges at the programme level. This can be judged as excellent if feedback is sought regularly from all stakeholders and used to make changes where required, good if it is sought frequently, adequate if occasional and poor if feedback is never gathered. To encourage collaboration despite contractual constraints, programmes can embed engagement expectations in TORs, allow flexible

commissioning, and reward responsiveness to user feedback. Regular touchpoints like inception meetings and joint reviews help build trust, while informal exchanges can surface needs that formal processes miss. These approaches promote shared ownership and more user-relevant outputs.

4.2 Justification

A central theme that emerged from the workshops and KIIs was the need to transition from transactional interactions to more collaborative engagement, fostering effective partnerships between research teams and stakeholders. Workshop discussions emphasised that this shift enhances research responsiveness, minimises inefficiencies caused by fragmented interactions, and ensures outputs are tailored to users' needs. Additionally, the KIIs reinforced the importance of prioritisation and adaptability - particularly for EF - where structured research selection processes and triage mechanisms help allocate resources efficiently while ensuring alignment with UK policy priorities and demand-driven research needs.

A key justification for having the criteria on technology was the growing reliance on technological innovation to enhance efficiency and maintain research integrity. Insights from both sources underscored the potential role of AI in plagiarism checks and content quality assessment, reducing manual review burdens while maintaining high standards. However, it was also discussed that protocols with respect to AI use must be developed, followed and maintained. Furthermore, optimal research administration was deemed crucial for agility in fund management, supporting ongoing adaptation to evolving priorities. Stakeholder relationships and strong supplier engagement were also included in the framework to emphasise proactive coordination and leveraging institutional knowledge - elements highlighted in the workshop as vital for efficient execution. Lastly, continuous learning mechanisms were identified as facilitating structured programme assessments and improvements, as well as strengthening knowledge-sharing and collaboration across research hubs in different geographies. Findings from both the KIIs and workshop suggest that efficiency is best achieved through strategic collaboration, informed decision-making, and adaptive learning.

It is also important to recognise that even if a project or programme performs excellently in terms of the efficiency with which it produces outputs, this may not automatically result in beneficial changes to policy, practice or strategy. This is because some of the factors which result in evidence being used are outside of the control of the researchers and/or research commissioners. While research advisers can increase the likelihood that the programme will lead to changes by efficiently identifying demand, external circumstance can prevent this. To ensure that an assessment of programme and portfolio management is not overly distorted by success or failure resulting from factors external to the programme, it is important to clearly distinguish the measures of efficiency presented here with those of effectiveness outlined in the following section.

4.3 Performance standards for efficiency

How well do funded projects produce high-quality, relevant outputs?			
Performance	**Adequate**	**Good**	**Excellent**
Sub-criteria	**Strategic relevance and structured prioritisation**: the programme has ad hoc or unclear selection processes to select high-impact, policy relevant research; weak or indirect link to policy priorities: research questions have room for improvement	**Strategic relevance and structured prioritisation**: the programme has structured processes to select high-impact, policy relevant research but are inconsistently applied; alignment with HMG priorities with few exceptions and logical research questions with minor exceptions.	**Strategic relevance and structured prioritisation: t**he programme clearly and consistently applies structured processes to select high-impact, policy-relevant research; strong alignment with HMG priorities and clear and logical research questions.
	Accessibility and coherence: findings are quite often overly technical and not very easy to understand.	**Accessibility and coherence**: findings are generally clear and well-structured with occasional technical ambiguities.	**Accessibility and coherence**: findings are consistently clear, well-structured, and easily understood by non-specialist audiences
	Programme outputs are of acceptable quality - timely, methodologically sound, and ethically grounded but with significant room for improvement	**Programme outputs are of good quality** - consistently timely, methodologically robust, and ethically grounded with minor exceptions	**Programme outputs are of high quality** - timely, methodologically robust, and ethically grounded
	Research design and delivery has very **limited collaboration** amongst all partners; outputs address general user needs but may lack specificity.	Research design and delivery has **some collaboration** amongst all partners; outputs address general user needs but may lack specificity.	Research design and delivery involve appropriate, **meaningful collaboration** among all partners; outputs align closely with user needs.
	Admin processes are manual, fragmented, or outdated; low technology use.	**Administrative processes** meet basic standards with some efficiency gains; technology use is moderate.	**Administrative processes** are streamlined and tech-enabled such as using dashboards; demonstrate adaptability and responsiveness.
	Weak partnerships: institutional knowledge is underutilized; little evidence of structured **learning or feedback loops**	Partnerships exist but are transactional; some **knowledge-sharing between research hubs occurs; absorptive capacity** is partial, some **data-driven improvements** are made, but lessons aren't always systematically shared	Robust partnerships (between users, research advisers, and suppliers) facilitate **mutual learning e.g. through learning and dissemination events; institutional knowledge** is actively used to inform decisions; stakeholders can act on research. **programme performance reviews** lead to documented improvements; lessons are shared across hubs systematically.
Performance will be rated as poor if any of the conditions of adequate are not met			

5 Aspects of value: Effectiveness

5.1 Potential criteria for effectiveness

Effectiveness is defined as: how well are the outputs used to inform policy, programming, diplomacy, partnerships and strategy?

Assessing the effectiveness of the EF and K4DD can be grounded in the existing FCDO framework for assessing research impact: instrumental use, conceptual use and process uses. These are grounded in well-established concepts of research impact[18], which can take a variety of forms.

A judgement on the overall standard of performance will require a combined consideration of the frequency and significance of impact from individual projects. Assuming that the potential impacts of projects are discrete, the assessment of programme effectiveness will require individual assessment of each project, or a representative sample thereof. Because of this, the various sub-criteria for effectiveness are described in relation to individual project-level outcomes. Developing a representative sample will require whoever is applying the framework to develop a set of key project characteristics that should be used to structure the sample, which could include, for example: geographic focus, policy sector, scale of project, research methods, or the type of delivery organisation (e.g. research university, NGO, consultancy)

- **Instrumental impact:** informing policies and programme delivery, influencing practice or services, shaping legislation, and changing behaviour.

In assessing instrumental impacts, the emphasis is on the extent to which the results of research have been used in informing policy or decision making with tangible results, either in the form of changes in policies and programmes, or in providing assurance that existing approaches are justified and effective. These changes could include changes to law or written policies or programme strategy, or alterations to ways of workings. For example, the instrumental impact may be judged as adequate if the evidence played some role in decisions on a policy, practice or diplomatic efforts. In other words, if the evidence had not been provided, the outcome of decision-making would have been different. That impact may be judged as good if that influence was substantial, meaning that the use of the evidence was critical to the decision that was made. Excellent instrumental impact would indicate that research evidence has led to very substantial / transformative change, such as the introduction of an entirely new policy framework, a massive scaling up of an intervention, or a new model of programme delivery.

- **Conceptual impact:** contributing to the understanding of policy issues and reframing debates

Conceptual impacts can be more challenging to identify. If evidence from the research is mentioned or referred in policy or strategy documents to improve understanding or reframe debates (for example, green papers in the UK government context[19]), that is a clear example of conceptual impacts. But often conceptual impacts are much less tangible, which also can make it challenging to judge the level of effectiveness. Research can be judged as adequate if it has informed users' thinking in some way, and so has the potential to

18 UK Research and Innovation / Economic and Social Research Council (ESRC) (2025) *'Defining impact', Impact Toolkit for Economic and Social Sciences* [online] Available at: Defining impact – UKRI

19 A Green Paper is a consultation document produced by the UK Parliament to allow people both inside and outside Parliament to give feedback on policy or legislative proposals

influence decision-making in the future. To be considered as good, the changes in thinking should be described as substantial, important or significant. Excellent conceptual impact would refer to circumstances where the change in thinking was described as very substantial or transformative and the nature of policy discussion changes substantially. Assessing conceptual impact will normally require qualitative feedback from evidence users to determine how significant the conceptual impact was r reported to be.

- **Process impact:** this includes impacts on capacity building (through technical and personal skill development) and connectivity (impacts on the existence and strength of networks of people and organisations who understand and can make use of the research)[20]

Process impact refers to wider impacts on the users of evidence beyond the specific content of that evidence. It can include improved general understanding of evidence and its value as well as strengthened relationships between decision-makers, collaborators, and researchers. The potential for process impact depends on the particular circumstances of a project, so it should not be expected that every project will have this type of impact, though it should be present across the portfolio as a whole, i.e. a good or excellent programme will take advantage of these opportunities when they arise. Making a judgement on the standard achieved for process use requires assessing the extent to which projects are improving perceptions of the value of research and/or the relationships between research advisers, research partners and users, such that additional value may arise from future collaborations and partnerships beyond the scope of the programme. It should consider the quality of relationships that enable evidence to be used meaningfully, whether enduring or short-lived. Evidence use is inherently relational, relying on trust, collaboration, and influence across teams and sectors

There is also a risk that research outputs may generate negative impacts among research users. If they are considered as not relevant, confusing, or incompatible with other objectives or values, users may decide to exclude research evidence from their decision-making and be less likely to consider research and evidence in future. While research evidence may often aim to change current thinking or practice, it must do so in a way which appears feasible and appreciates the context in which decisions are made.

The threshold for adequate is set as not producing these negative outcomes, and where users feel that outputs have improved their understanding of evidence and are more likely to consult evidence in future decisions. A process impact would be considered good when a user's capacity to understand research and apply it appropriately to decision-making has been clearly improved and/or there are indications that researchers and users are more likely to work together in the future. While relevance remains essential, it must allow for 'constructive disruption' where research enables users to rethink, reframe, or evolve their decision-making approaches. Therefore, process use should be assessed not only by improved technical understanding, but also by a user's increased confidence and openness to engage with unfamiliar or challenging evidence, even when it contradicts current norms.

To meet an excellent standard, research users should report that their understanding of how evidence is used, and the value they place on using evidence, has been very substantially enhanced or transformed. Alternatively, or in addition, there should be substantive follow-on activity, for example, a user commissioning additional work from a research partner or consulting them on other decisions. Platforms like K4DD may catalyse impactful connections during critical moments, such as crises, by convening diverse actors and integrating multiple ways of knowing. While long-term partnerships are ideal, short-

20 Shaxson, L. (2016) *Achieving Policy Impact: Guidance Note (DEGRP)* [online]. Available at: DEGRP-Impact-guidance-note.pdf

term collaborations can also yield significant conceptual or instrumental impact. Evaluations should therefore recognise temporal dimensions of connectivity and account for external factors, such as organisational restructuring, that may limit the visibility of longer-term outcomes.

In addition to the criteria described above, existing FCDO impact frameworks also refer to symbolic use, where evidence is deployed to lend legitimacy to decisions already preferred. This can be a legitimate use of evidence and can be instrumental or conceptual. However, there is also a risk that evidence will be used selectively, or 'cherry-picked' to justify a decision in a way that does not reflect the evidence in its entirety. When assessing effectiveness, users of this framework should be mindful of this risk. Instances where evidence is used symbolically in way that misrepresents, or selectively applied evidence should be considered as inadequate instrumental or conceptual use.

The extent to which these criteria are achieved will vary between projects in the programmes, with some projects leading more naturally to instrumental impacts and others leading to more conceptual impact. Furthermore, it is crucial to stress that the impact of research and evidence on decision-making should be proportionate and appropriate to the robustness and rigour of the findings. This has different implications for the two programmes. As K4DD outputs are based on the synthesis of existing bodies of evidence, they are to some extent more robust than individual primary research projects. However, as K4DD outputs are produced relatively rapidly, one may question whether their findings are comprehensive enough to inform major decisions. Conversely, EF outputs are produced over a longer timeframe, providing more opportunity to ensure their internal validity. However, making decisions on the basis of a single primary research project, that has not been reviewed or compared with other results by the wider research community, carries the risk that the findings may be outliers or have misinterpreted the data. Assessments of the appropriateness of research use must therefore be made on a case-by-case basis. A programme level assessment should be informed by how systematically the programme highlights and manages these questions, for example, by ensuring all outputs include references to the limitations of their findings. Projects should systematically note whether they align with previous, similar studies, and be transparent about the level of certainty of their results, or whether further research is needed to resolve outstanding questions that may impact on how findings should be used. It is important that the assessment of impact considers the overall effect of programme outputs, rather than focussing exclusively on the impact of single outputs. This could be done with reviewing a sample of outputs as well as assessing guidance on the production of outputs. Do outputs consistently refer to the level of certainty of their results and note priorities for further work to improve the strength of evidence, where warranted? Particularly for K4DD, the best impact may arise from a body of help desk inquiries and some learning services which collectively support shifts in attitudes or policy and programming.

To maximise value for money, we expect that the portfolios of the EF and K4DD will achieve a balanced spread across these impact types, ensuring that investments are strategically distributed to address conceptual impacts, instrumental impacts and process impacts. As project effectiveness involves circumstances outside of the control of researcher partners or programme managers such as budget constraints, power dynamics, shifts in policies, it is also unrealistic to expect all projects in a programme to achieve good or excellent standards for effectiveness. EF and K4DD may not control downstream funding decisions, but their role in shaping strategic thinking, fostering collaboration, and generating high-quality evidence remains vital. Even if a programme investment is later deprioritised or defunded, the influence of EF and K4DD on how decisions were framed, what evidence was considered, and how stakeholders engaged still constitute meaningful outcomes. Their effectiveness should

therefore be assessed not only by the fate of individual programmes, but by their contribution to a more evidence-informed ecosystem as well as the overall effect of the full portfolio of research outputs. To make systematic and transparent determinations of overall programme effectiveness, programme managers could construct a matrix to specify programme level expectations of the percentage of projects which meet the defined standards. An illustrative example is given in Table 4. However, it is important not to apply such an approach mechanistically, since the effectiveness of the programmes as a whole is a function not only of the proportion of outputs have impact, but also on the extent of that impact, as the rubric makes clear. A smaller fraction of projects having a profound influence on important policy areas or large programmes may be judged to compensate for others that have relatively little impact.

Table 4: Relationship between aggregate research project effectiveness and programme level assessment (illustrative)

Programme rating	Percentage of projects reporting impact on a given sub-criterion/impact dimension
Excellent	'Very often'- more than four cases in five
Good	'More often than not'- e.g. more than half of cases, but less than four in five
Adequate	'Sometimes' e.g. less than half of cases, but more than one in five
Poor	'Rarely'- fewer than one case in five

5.2 Justification

The typology of instrumental, conceptual and process impact outlined above is established by FCDO. It builds on widely used framework for describing impact, notably by UK Research & Innovation (UKRI) and incorporates additional categories such as connectivity (incorporated here into process use), which was introduced by Overseas Development Institute in a paper commissioned to provide an impact framework specifically for development contexts.

Instrumental impacts are often prioritised in impact assessments, as they are frequently simpler to measure. However, achieving a clear instrumental impact can rely on the alignment of factors outside of the control of the researchers. Furthermore, reviews of research impact have shown that conceptual impacts are often a necessary precursor to instrumental impacts[21]. Decision-makers need to understand the value of research findings or evidence more broadly, before incorporating them into specific policy decisions. For example, a previous evaluation of research impact concluded:

21 France, J., Rajania, A., Goodman, R., Ram, M., Longhurst, R., Pelka, V. & Erskine, C. (2016) *Evaluating the Impact of the ESRC-DFID Joint Fund for Poverty Alleviation Research: Final Report* [online]. Available at: https://www.ukri.org/wp-content/uploads/2022/08/ESRC-080822-EvaluatingImpactJointFundPovertyAlleviationResearch-FinalReport.pdf

> "Most evidently, Conceptual Impacts and Capacity-building[22] may be inextricable; in addition, for example, Instrumental Impacts may rest upon Capacity-building and/or be the manifestation of Conceptual Impacts."[23]

In the context of many research programmes, capacity strengthening often refers to improving the capabilities of individual researchers (often early-career researchers) or strengthening the capacity of research organisations.[24] In the context of assessing the effectiveness of the EF and K4DD, the focus will more often be on the capacity of decision-makers i.e. users to effectively use research evidence. Capacity here can be understood to include several things, including the technical skills and knowledge to understand evidence, the appreciation of its value to decision-making, and the ability to incorporate evidence into policy and practice.

Likewise, connectivity can refer to linkages among groups of researchers (particularly across different disciplines). This is likely to be of some significance for the EF, where different disciplinary approaches must be combined to address a practical question, or in K4DD, where evidence is synthesised from multiple disciplines. However, as cross-disciplinary working can be a long-term process, in many cases we expect the most relevant connectivity to be in terms of building trust and collaboration between research partners and users or decision makers, where greater value for money is produced by stimulating relationships that extend beyond the context of a particular project. For example, decision makers may consult researchers on future policy questions.

[22] Included within process impact in this framework

23 Denyer, D. & Meagher, L. R. (2013*) Research impact on practice: case study analysis* [online], p. 33. Available at: Research Impact on Practice: Case Study Analysis

24 The development of capacity in research suppliers from the Global South is addressed under equity.

5.3 Performance standards for effectiveness

How well are the outputs used to inform policy, programming, diplomacy, partnerships and strategy?			
Performance	**Adequate**	**Good**	**Excellent**
Sub-criteria	The programme demonstrates some **instrumental** impacts in that it has informed policy decisions, programme design or implementation.	The programme demonstrates substantial **instrumental** impacts which have significantly influenced policy decisions, programme design or implementation.	The programme demonstrates very substantial **instrumental** impacts which have transformed policy, programme design or implementation.
	The programme demonstrates some **conceptual** impacts which have informed strategy documents or how decision makers think about the relevant policy, practice and implementation challenges.	The programme demonstrates substantial **conceptual** impacts which have significantly influenced strategy documents or how users think about the relevant policy, practice and implementation challenges.	The programme demonstrates very substantial **conceptual** impacts which have inspired new strategic thinking, substantially reframed policy debates or transformed users' thinking about the relevant policy, practice or implementation challenges.
	Process impact: Research outputs have had some positive impact on users' ability to understand and use research and evidence, their view of the value of research, and the likelihood that will use research to inform decision-making in the future. They are somewhat more likely to seek out evidence and engage with researchers to inform future decisions.	**Process Impact:** Research outputs have substantially improved users' ability to understand and use research and evidence, their view of the value of research, and the likelihood that will use research to inform decision-making in the future. They are substantially more likely to seek out evidence and engage with researchers to inform future decisions.	**Process Impact:** In terms of capacity building, research outputs have very substantially improved users' ability to understand and use research and evidence, their view of the value of research, and the likelihood that will use research to inform decision-making in the future. The programme has very substantially strengthened **connectivity** and trust between researchers and users and generated significant ongoing/future collaboration or partnership.
Performance will be rated as poor if any of the conditions of adequate are not met			

6 Aspects of value: Equity

Here, equity is defined as: do the processes for prioritising, commissioning, conducting and using research involve a sufficient diversity of experiences, voices, and perspectives, including those that are traditionally marginalised? Are the benefits of the research fairly distributed, and do they meaningfully reach and represent underserved or disadvantaged groups?

6.1 Potential criteria for equity

- **Catering to a diverse user base:** The programme is visible to and utilised by a wide range of users within FCDO regions (post and central teams) and functions (i.e. across different directorates, ODA and non-ODA) and other relevant UK govt departments, as well as used by non-UK government partners and stakeholders where appropriate, and demand is managed in a way which ensures that usage is not concentrated in small pockets. This may be tracked by examining diversity in the origin of requests received under the programme, basing the judgement on the proportion of directorates and regions submitting requests each year. For example, if all requests come from 2-4 directorates or regions, the programme may be considered 'Poor' on this criterion, whereas if requests originate from nearly all directorates and/or countries within the regions, the programme can be deemed 'Excellent' on this criterion.
- **Diversity within outputs commissioned:** This criterion subsumes three aspects:
  - The subject of research outputs commissioned under the programmes cover a range of thematic and/or geographic areas, provided they are in line with the UK geographical and thematic priorities. Judgement on this criterion can be based on an annual mapping of projects commissioned across priority areas (e.g. climate change, science and technology, humanitarian response etc.). If projects commissioned cluster within a few topic areas or regions, the programme may be deemed 'Poor' whereas a high degree of diversity would warrant an 'Excellent' judgement.
  - Considerations of equity and intersectionality (such as gender, disability, and socio-economic status) are embedded within research outputs as far as feasible, leading to improving outcomes. A review of the finalised ToRs and the content of the outputs (or a representative sample of these) can be used to track if (and how) intersectionality is addressed (e.g. through disaggregated findings, equity analysis). For instance, say if 75% or more outputs commissioned embed equity considerations meaningfully, then the programme can be deemed 'Excellent' on this criterion.
  - The extent to which the outputs include, in an appropriate manner, knowledge and perspectives beyond dominant academic sources - such as grey literature, local knowledge captured through FGDs and KIIs with local actors, and perspectives from non-Western and/or marginalised groups. This could be gathered through a review of the sources which an output draws upon, as well as peer review feedback. If there is clear, meaningful integration of non-dominant knowledge sources in most outputs (or a representative sample of these), then the programme may be deemed 'Excellent' whereas if they are rarely or never included in any outputs, a 'Poor' judgement would be appropriate.
- **Diversity of research partnerships and commissioning:** The programmes engage local or, regional, and underrepresented research partners (particularly for research on the Global South) to build technical capacity and improve contextual relevance. This can be judged based on the share of commissions (measured as share of budget and/or number of commissioned studies) going to regional or local institutions and/or locally

based PIs, as well as some evidence of capacity building activities for local/regional suppliers (e.g. market engagement events, joint authorships of outputs). Feedback from research partners commissioned (on their roles and experience with knowledge exchange and learning) can also be utilised.

- **Accessibility of research outputs:** This in turn, covers three aspects:
  - Where suitable, outputs commissioned by the programme are open-source and publicly available, contributing to the global knowledge commons. Judgement for this criterion can simply be based on the number of (ODA) studies or outputs published on the gov.uk website and/or the Evidence Fund and K4DD websites. If most studies produced remain unavailable for public use, the programme may be deemed 'Poor' on this criterion.
  - Outputs are presented in a format that are appropriate for use understanding of the primary users of the research (which may include non-technical audiences). Based on user needs and/or requests, formats may range from technical reports, to presentations, infographics, or policy briefs or multi-lingual outputs. The requested output format(s) used, and how satisfactory users found them, may be tracked through feedback from users.
  - Outputs are published in open, editable formats like .odt, to support users of adaptive technologies (e.g. screen readers, text-to-speech tools). A review or audit of the formats of all published outputs, or a representative sample, can be used for judgement on this criterion.

As assessment of equity cuts across the entire results framework of a programme (see **Figure 1**, above) it is important to also consider equity in terms of the outcomes and impact which emerge from programme activities. However, it is important to make these assessments alongside other measures of the significance of evidence outcomes. For this reason, equity is also included below in the criteria for assessing overall cost-effectiveness.

6.2 Justification

The aforementioned criteria for equity in the value-for-money (VfM) framework is based on the workshops and key informant interviews (KIIs) and closely guided by a review of the programmatic documentation. The criteria pertaining to the diversity within outputs commissioned – especially in terms of thematic and/or geographic areas and on considerations of equity being embedded within research outputs – are outlined in the original business cases for both programmes. Similarly, achieving diversity of research partnerships by engaging underrepresented researchers or supplier firms has been a key component of the programme's original business case. This is also a key focus area for stakeholders whose roles involve commissioning research and evidence synthesis and engagement with suppliers, as the KIIs reveal. The criterion pertaining to the diversity of the programmes' user base reflects an addendum to the business case for the Evidence Fund (to expand non-ODA research), as well as concerns raised by several stakeholders (during the KIIs) on the awareness and/or use of the programmes potentially being limited to certain teams / directorates of FCDO - as opposed to the broad user base envisioned. Lastly, the criteria on the accessibility of research outputs have been included on the basis of feedback provided by programme users.

While the criteria above defines how equity may be assessed across both demand-driven research programmes, it is important to acknowledge the differing strategic objectives and operating models of Evidence Fund and K4DD when applying the framework. As a

consequence of this, performance on certain equity dimensions may legitimately vary between the two programmes, reflecting a trade-off between equity, speed, scope, and strategic focus. The criteria above should therefore be interpreted in a context-sensitive manner when applied to each programme, as highlighted in Box 1 below, and an assessment of the two programmes is likely to show differences between them in performance across the full set of VfM criteria, reflecting these different priorities.

6.3 Performance standards for equity

Do the processes for prioritising, commissioning, conducting and using research involve a sufficient diversity of experiences, voices, and perspectives, including those that are traditionally marginalised? Are the benefits of the research fairly distributed, and do they meaningfully reach and represent underserved or disadvantaged groups?

Performance	Adequate	Good	Excellent
Sub-criteria	**Catering to a diverse user base:** the programme's user base is concentrated in a few teams or stakeholder groups; ad- hoc efforts to improve awareness or outreach exist but lack any strategic targeting.	**Catering to a diverse user base:** the programme reaches a moderately diverse user base across some FCDO teams and external partners; efforts to improve visibility or outreach are evident; usage of the programme is not overly concentrated in certain teams or stakeholder groups, although some imbalances exist.	**Catering to a diverse user base:** the programme has a broad and diverse user base across FCDO regions, themes, and functions (different directorates, ODA and non-ODA) along with external partners. Active engagement or outreach mechanisms to increase awareness of the programme and increase the equity of usage are in place, and demand is proactively managed to ensure balance across user types, within an overall prioritisation framework.
	Diversity within outputs commissioned: thematic and/or geographic diversity is somewhat limited in the research outputs commissioned; the inclusion of equity considerations in outputs is inconsistent; outputs mostly draw upon conventional academic sources alone.	**Diversity within outputs commissioned:** outputs cover a reasonably broad array of subjects and/or regions (in line with FCDO priorities), equity considerations are generally addressed within outputs wherever relevant; there is some or occasional use of non-traditional knowledge sources or perspectives.	**Diversity within outputs commissioned: r**esearch outputs consistently span diverse themes and geographies (in line with the FCDO's strategic priorities); equity and intersectional issues are meaningfully integrated in outputs wherever appropriate; outputs regularly draw on sources such as grey literature, local knowledge and underrepresented perspectives.

	Diversity of research partnerships and commissioning: local or regional partners are occasionally engaged, usually in a limited capacity (e.g. a sub-contractor for a small workstream, on very few contracts); there is some (very limited) investment in capacity-building for such suppliers; and heavy reliance on UK-based or other Global North partners.	**Diversity of research partnerships and commissioning:** regional or local partners are involved in many outputs; there are some capacity-building efforts and evidence of inclusive partnerships (even where UK or Global North organisations are leading delivery).	**Diversity of research partnerships and commissioning:** Local or regional research partners (firms and/or individuals) are consistently engaged as primary or co-leads for commissioned outputs; there is demonstrable evidence of investments in building the long-term capacity of research partners from underrepresented groups and/or geographies; representation from underrepresented groups and Global South-based research partners is high.
	Accessibility of research outputs: outputs are available and accessible to (most) internal audiences, where appropriate, but not routinely shared in the public domain even when appropriate; there is little effort made to adapt content or formats for wider accessibility, including for use with adaptive technologies; Most users report the outputs being difficult to understand and engage with.	**Accessibility of research outputs:** most outputs are available in formats supported by adaptive technologies, and there is evidence of dissemination beyond the commissioners / primary users of outputs; users report the outputs being easy to engage with and understand (with some exceptions); most outputs are available online (where appropriate).	**Accessibility of research outputs:** outputs are consistently made open access (where appropriate) and available in formats supported by adaptive technologies; users (including non-technical users) report the outputs being easy to engage with and in line with their needs
Performance will be rated as poor if any of the conditions of adequate are not met.			

7 Aspects of value: Cost-effectiveness

Cost-effectiveness is defined as: what is the ultimate final impact and value of the evidence generated by the programmes on development priorities, the UK's diplomatic goals and the culture of evidence use within FCDO and its partners, and does it justify the inputs that are invested in it?

7.1 Potential criteria for cost-effectiveness

We assess cost-effectiveness on the basis of stakeholder value - i.e., to what extent do the programmes meet their value propositions as expressed in sections 2.1, 2.2**Error! Reference source not found.Error! Reference source not found.** and 2.3**Error! Reference source not found.**. The key differentiating factor is to anchor the cost-effectiveness assessment in the value stakeholders place on the programmes as distinct from measurable outputs or outcomes.

To achieve this, the cost-effectiveness assessment must go beyond the consideration of the type of influence on decision-making, or furthering UK's foreign policy objectives through diplomacy or strategic partnerships, to take account of the potential impact on the lives of ultimate beneficiaries, both in FCDO countries of operation and in the UK. While it will often be the case that these impacts will only be realised outside the scope of project timelines, in most cases an assessment of the potential scale of impact should be possible (e.g. is the research influencing a national policy that will potentially affect millions vs. a very localised change). Evidence of policy influence alone is insufficient to provide assurance of value for money, as the extent of that influence may be insufficient to justify the cost involved. Of course, when considering the potential impact on beneficiaries, questions over the specific contribution of research and evidence to that change will arise. Even when there is uncertainty over this contribution, the scale of potential change remains an import consideration in assessing overall value.

Our framework therefore proposes that cost-effectiveness is assessed in two sets of criteria, the first of which refers to the overall perceived value to stakeholders. Unlike outcomes and impacts, which are real changes in people, groups and behaviours caused by the research programmes, stakeholder value represents the meaning and importance that different groups assign to the programmes. Value is not inherent in activities and results - it is actively placed on them by stakeholders according to their needs, preferences, and priorities.

- **Overall stakeholder value**: the programmes deliver on their promise to users, providing value as defined by the perceptions, experiences, and judgements of those who engage with or benefit from the programmes. This refers not just to outcomes or impacts achieved but to the merit, worth, and significance that stakeholders attribute to the programmes, including their satisfaction, relevance, and perceived benefits *relative to costs*. For example, stakeholders may place high value on a project that fostered meaningful collaboration or addressed urgent concerns, even if the measured policy outcomes were less significant. This underscores the importance of understanding stakeholder perspectives—not just as a measure of satisfaction, but as a critical lens through which the relevance and impact of programme outcomes are interpreted. Overall stakeholder value is therefore assessed by how worthwhile and significant stakeholders find the programmes, *relative to resources used*, as expressed directly by them - such as through feedback, ongoing engagement, championing, or continued

investment. [25] A programme can be considered **excellent** when stakeholders clearly see it as valuable and relevant as reflected through positive feedback and/or active engagement, It is rated **good** if most stakeholders find it useful, though there may be moderate engagement and limited investment. A programme can be seen as **adequate** when stakeholder views are mixed, with only modest signs of value. The frequency-based ranges suggested in Table 4 could also be considered for this sub-criterion.

- **Uniqueness**: the programmes meet real information needs addressing global or internal evidence gaps and have unique strengths that make them distinct and valuable relative to alternatives. The differences could be in content, delivery, outcomes or method. For example, one programme may stand out by producing rapid-turnaround evidence tailored to policymaker timelines, while another may be valued for its participatory research methods that build local ownership. These distinctive features enhance their value and relevance to stakeholders, and distinguish them for alternatives, making them a preferred source of evidence and insight. A programme can be considered **excellent** if it clearly addresses a defined evidence gap in a unique and valuable way, **good** if it responds to relevant needs and shows some distinctive features and **adequate** if it meets basic information needs and offers limited differentiation. This could be assessed qualitatively, through KIIs with key, selected stakeholders, or could be addressed through a user survey and then make use of the frequency-based ranges suggested in Table 4.

The second set of criteria seeks to estimate the significance of the impacts identified in the effectiveness assessment, in terms of their **ultimate impact on the final beneficiaries of decisions made**. This is a programme-level assessment i.e. it should consider the aggregate effect of the entire programme. However, it will need to be assessed by aggregating information on the impact of individual research projects.[26] Impact should consider several dimensions:

- **Scope:** the number of individuals and/or geographic extent of the changes that may result from research (e.g. national, regional, individuals or communities). A project can be rated **excellent** if its impact reaches a large number of individuals or spans a wide geographic area – for example, contributing to decisions that improve the lives of millions or tens of millions of individuals.
- **Depth:** the scale of the change on the lives of those affected. A project can be considered **excellent** if it leads to substantial changes in the lives of beneficiaries such as helping to bring about substantially improved access to essential services.
- **Equity:** the extent to which benefits will be distributed equitably, or benefit marginalised or hard to reach populations. A project may be considered excellent if it contributes to a substantial improvement in the lives of disadvantaged or marginalised groups – such as helping to improve school attendance or learning in groups that have been underperforming.

Note that these dimensions are complementary – for example, a project may have a smaller scope but this may be compensated by a substantial contribution to equity. The assessment

[25] Feedback on stakeholder value can typically be provided by research customers and technical advisers. Research customers focus primarily on how effectively the programme met their needs, however they define them, while technical advisers may consider both the quality, effectiveness and cost of outputs. Their insights gathered through engagement, reviews, or continued support help determine how the overall judgement about how worthwhile and significant a programme is perceived to be relative to the resources invested. It would also possible, if desired, to give other stakeholders information on the costs of the projects whose outputs they used and ask them to make their own assessment of value created relative to cost.

[26] See Annex D for a discussion on ways in which project-level data can inform the programme-level assessment. It may be based on either all projects, if relevant data is available for, or for a sample.

of scope and depth should be interpreted in relation to what decisions or actions the project seeks to influence. For example, a national-level study on agricultural policy might influence guidelines affecting thousands of farmers across multiple regions demonstrating wide scope but only result in modest changes to farming practices (moderate depth). In contrast, a local pilot programme that works intensively with a small group of farmers may lead to transformative improvements in income, resilience, and sustainability—showing deep impact but limited reach.

Individual projects can be assessed for the extent of their impact and this can be compared to their cost. The comparison must take into account the wide variation of final impacts (different sectors, types of interventions, geographies etc) and the wide variation in effectiveness that is possible for different projects. At an individual project level, measures of final impact (such as increased numbers of children enrolled in school or lives saved) could be presented as a cost-effectiveness ratio. However, it may be the case that projects have affected decisions which have a range of outcomes (e.g. on both school enrolment and child labour) which cannot be summarised into a single outcome measure. Furthermore, there would be no single natural outcome measure that could summarise final impacts across the wide variety of areas that may be influenced by the research programmes. Converting all of these into a money-metric value to undertake a cost-benefit analysis is likely to be an excessively demanding task and very sensitive to a large number of assumptions that would need to be made. For that reason, we recommend that costs and outcomes of various types for each project be compared more qualitatively, with the assessment making use of the rubric to make an overall judgement on impact vs cost. This approach (which is an extension of a cost-consequence analysis giving an initial view of the landscape before undertaking CEA) encourages an explicit recognition of multiple possible outcomes (including negative ones) and trade-offs. This information can then be brought together (in a listing/tabular format, not an averaging process) across relevant projects (either all, or a sample) to make the programme level assessment. A relatively small number of projects that are able to demonstrate a very large impact on final beneficiaries may, rightly, drive a positive overall assessment of the programme as a whole. The addition of a rubric is profound because it extends CCA into something that can be used evaluatively.

7.2 Justification

Attributing changes to a particular research output or measuring its contribution can be a complex and resource intensive process, involving detailed evaluation methods (e.g. process tracing, outcome mapping, etc.). For the purposes of this VfM framework, a simpler process of seeking feedback from decision makers and making a judgement about the plausibility of contribution claims may be appropriate.

Quantifying contribution within a Value for Money (VfM) framework does not require complex attribution methods; instead, it can be approached through structured, light-touch indicators. By gathering feedback from decision-makers through surveys or semi-structured interviews and assessing the plausibility of influence based on timing, relevance, and engagement, evaluators can assign contribution scores or ratings. Additional signals like citations in policy documents, follow-on activities, or increased collaboration can be tracked to strengthen the case.

The assessment of the significance of impact needs to be nuanced and context specific. There may be inherent trade-offs between scope and depth that need to be made on a case-by-case basis in judging the overall cost-effectiveness of the influence on decision-making. Also, in general terms, a project which influences nation-wide policy change can be considered as more effective than one that only affects a small population. However, if that population

represents a marginalised group (e.g. those in extreme poverty, people with disabilities or disadvantaged ethnic minorities) then this should be considered when judging overall effectiveness.

Cost-effectiveness criteria can be used to assess individual projects or the total value of a programme. Performance will vary between projects, and some projects may demonstrate poor cost-effectiveness due to external factors, despite being well-designed and implemented. Having a few projects with poor cost-effectiveness does not necessarily mean that the programme overall should be judged as poor, and doing so may create a risk averse culture that limits the value of the programme as a whole. In developing a summary assessment of a programme's cost-effectiveness, users of this framework should identify their appetite for risk across the portfolio. Some programmes may be high-risk, high-reward- they could fail, but if they succeed, the impact is huge. Others may be low-risk, low reward- more predictable, but with modest outcomes. The approach outlined above is intended to allow for this variation.

As the assessment of cost-effectiveness spans the entirety of the results chain, at this stage it is important to carefully consider the relationships between the assessments of other criteria and sub-criteria. Reasonable trade-offs may need to be made between various criteria, and weaknesses in delivery of some may be managed by subsequent programme activities. For example, a project may not produce timely or accessible outputs, but adaptive management of the programme could counteract these factors to achieve policy impact.

Other criteria represent critical points of failure. For example, though rare, if the quality of a research project is poor (meaning that the validity of its findings is questionable) then this would overshadow any other measures of the project's effectiveness. If such work achieves substantial policy impact, then this would be a negative impact.

A balanced assessment of the dimensions of significance is also required. Research which leads to impact on a wide scope, but with only moderate depth, may or may not be considered more cost-effective than impact which is transformative to a smaller population. The subjective assessment of stakeholders, noted at the beginning of this section, is therefore essential in making these judgements.

7.3 Performance standards for cost-effectiveness

What is the ultimate final impact and value of the evidence generated by the programmes on development priorities, the UK's diplomatic goals and the culture of evidence use within FCDO and its partners, and does it justify the inputs that are invested in it?			
Performance	**Adequate**	**Good**	**Excellent**
Sub-criteria	**Overall stakeholder value:** the programmes meet some stakeholder expectations, but value is mixed or limited to certain groups.	**Overall stakeholder value:** the programmes largely meet stakeholder expectations, with positive value reported by most users.	**Overall stakeholder value:** the programmes fully deliver on their promise to users, with stakeholders reporting high satisfaction and clear, meaningful value as they define it.
	Uniqueness: the programmes meet some information needs but lack clear differentiation or unique value.	**Uniqueness:** the programmes address most key needs and have distinctive strengths compared to alternatives.	**Uniqueness:** the programmes consistently meet priority information needs and offer unique, highly valued strengths compared to alternatives.
	The instrumental and conceptual impacts of the programme may have some limitations in **scope**, be only incremental in **depth** or fail to comprehensively promote **equity**, though it is still clear that likely benefits exceed the cost of programme delivery.	The instrumental and conceptual impacts of the programme have potential to improve the lives of ultimate beneficiaries in ways that exceed the cost of the programme substantially, through being either wide in **scope**, significant in **depth** or promoting **equity** by reaching marginalised populations.	The instrumental and conceptual impacts of the programme have strong potential to improve the lives of ultimate beneficiaries in ways that exceed the cost of the programme by orders of magnitude, through being either extensive in **scope**, transformative in **depth** or promoting **equity** by reaching highly marginalised populations.
Performance will be rated as poor if any of the conditions of adequate are not met.			

8 Recommendations on data collection and application of the framework

To develop this VfM framework, we followed steps 1-4 of the process outlined in Figure 2 (Section 1.1). In the following paragraphs, we provide guidance for those who will use the framework, following steps 5-8: gather evidence, analyse evidence, synthesise evidence using the rubrics to make explicit VfM judgements, and report findings. We also provide additional guidance on application of the framework including differences between EF and K4DD, data sources, and potential application of the approach to other programmes.

8.1 Step 5: Gathering evidence

The VfM assessment starts with collecting the right evidence for each criterion and sub-criterion. This evidence may be both quantitative (e.g., cost centre reports, output indicators) and qualitative (e.g., interviews, surveys, case studies). It is important to use multiple sources to provide a comprehensive view and to triangulate findings. Where possible, data collection should align with existing MEL systems to avoid duplication and make use of established processes. Specific evidence sources suggested are detailed in section 8.6 below and further details on suggested approaches are given in Annex D.

8.2 Step 6: Analysing evidence

At this stage, each stream of evidence is analysed independently to generate findings that address specific criteria and sub-criteria. For each VfM dimension, the aim is to generate discrete, well-supported pieces of evidence (e.g., on commissioning speed, diversity of users, the nature of partnerships, etc.) that are linked directly to relevant rubric descriptors. This analysis should be conducted by team members with a good understanding of the evidence context, and the process should include checks for bias, consistency, and reliability. Where appropriate, create concise summaries or tables capturing key findings for each criterion, ensuring the evidence for each stream is as clear and objective as possible. A series of annexes may be produced for the report, each focusing on a separate stream of evidence (e.g., cost data, survey feedback, documents analysis, etc.).

8.3 Step 7: Synthesis and judgement

Once each evidence stream has been analysed, the next step is to bring all findings together and systematically consider them against the criteria and standards. While preliminary judgements may be reached by an evaluator (or ideally two or more evaluators working collaboratively), we recommend that these preliminary judgements be discussed and validated through a collective process, such as a panel or facilitated group session, where all relevant evidence is reviewed, discussed, and weighed. Participants should represent a broad range of perspectives, including programme managers, MEL specialists, end users, and ideally, one or more independent reviewers.

For each criterion and sub-criterion, the group compares the findings from the evidence with the rubric's descriptors for 'adequate', 'good', and 'excellent' performance. Through structured discussion, the group seeks to reach consensus on which standard the evidence best aligns with, noting any uncertainties or dissenting views. The judgement process should be well

documented, referencing the evidence used, the rationale for each conclusion, and any agreed recommendations or insights for improvement. This transparent synthesis ensures VfM ratings are robust, credible, and clearly understood by stakeholders, which increases the likelihood that findings will be used to inform programme improvements.

It is important to approach the process as evaluative reasoning rather than mechanical scoring. Good practice involves:

- Familiarising those involved with the general approach (as set out in OPM's Guide), and the specific standards and evidence used in this assessment
- Being clear that deliberation, debate, and uncertainty are expected - full consensus may not always be possible, but a fair and transparent process builds credibility
- Using the process as an opportunity for open enquiry and learning - for example, inviting discussion on why performance met or did not meet a standard, and what could be done differently
- Anchoring judgements in evidence, guided by the rubrics, while exercising contextual professional judgement: rubrics don't make judgements - people do.

8.4 Step 8: Reporting and using the results

The final step is reporting findings in a way that is clear, balanced, and actionable. The report should address each criterion under the 5Es separately, highlighting overall ratings, strengths and areas for improvement, and practical recommendations. Where appropriate, ratings may be summarised in a scorecard or similar, together with explanations and context. A key principle at this stage is "show your working"; the purpose of reporting is not just to pass judgement, but to present a defensible argument for the judgements, based on the evidence, criteria, and standards, and to foster learning, adaptation, and ongoing improvement.

In practice, reports often include a summary table of ratings and rationale for each criterion in the executive summary, a succinct performance story structured around the criteria which presents the key pieces of evidence supporting each rating, and a detailed series of annexes providing extra details, such as methods and analysis of each evidence source.

Those seeking further guidance on this approach should refer to OPM's Guide.[27]

8.5 Application of the framework: key differences between Evidence Fund and K4DD

As noted in the introduction, the framework was designed to inform VfM assessments at the programme level, that is, to assess each programmes as a whole. It was designed to be broad and flexible enough to be used with the EF or K4DD, which means that some individual elements may be less applicable to any particular programme. For example, EF employs a strong prioritisation process to ensure that its more expensive outputs are produced only for the issues of most strategic importance to the organisation. K4DD offers its, much cheaper, services to all comers on a first-come, first-served basis. The two programmes will therefore be rated differently on sub-criteria for effectiveness and equity, with EF rated better on

27 King, J., Wate, D., Namukasa, E., Hurrell, A., Hansford, F., Ward, P., Faramarzifar, S. (2023). *Assessing Value for Money: the Oxford Policy Management Approach*. Second Edition. Oxford Policy Management Ltd.

‘strategic relevance and structured prioritisation’ (part of effectiveness) but lower on ‘catering for a diverse user base’ (part of equity).

By not limiting the framework to sub-criteria on which both programmes place similar priority, it provides a basis for a high-level comparison between the two across all five criteria. Box 1 below brings out some of the areas in which we would expect a priori differences between the two programmes, given their different strategies and budgets.

Box 1: Expected differences in the application of the VfM framework between the two programmes

Criteria	Sub-criteria	Key differences
Economy	**Administrative and governance structures are streamlined**	Studies under K4DD are as short as 5.5-6 days of work while EF outputs have a timeline of at least 3-6 months, therefore the streamlining of the admin and governance structures will be based on the context of the study/project.
	Timely and streamlined commissioning process	The commissioning process will vary between K4DD and EF. K4DD will has a light-touch process as it focuses on small-scale rapid studies through a fixed delivery consortium, while larger projects through EF follow a more rigorous path through competitive tendering.
	Operational costs	These would be on the higher side for EF and lower for K4DD. The judgement must be made by considering that the quality and scale of the outputs will be different between the two programmes.
	Transparent fund management	Transparent budgeting of input days is essential, as misalignment with methods can distort costs. EF has more standardised procedures for defining input days in relation to study objectives and methods and emphasises methodological clarity. K4DD shows variability in how input days are defined and allocated.
	Complementarity	While both EF and K4DD aim to leverage existing networks, EF does so more systematically, minimising duplication specially as it is more long term. K4DD shows occasional overlap and fragmented reporting
Efficiency	**Strategic relevance, structured prioritisation and getting the right research questions**	EF has a strong prioritisation process, focussing its larger resources on the most strategic areas for FCDO. In contract, K4DD does not have a prioritisation process. The rapid reviews or synthesis papers are approved on first come first serve basis. K4DD would therefore be expected to score less well on this criterion, reflecting different strategies.
	Accessibility and Coherence	Across both programmes**,** EF and K4DD, clarity and coherence, such that their logic and implications are understandable to non-technical audiences should be a key criterion for judgement
	Quality Outputs	K4DD focuses on rapid, high-level synthesis whereas EF focuses on in-depth analysis, which is much better resourced. K4DD would be expected to score better on timeliness, while methodological rigour will be higher for EF outputs. Both should expect to meet programme-specific quality expectations.
	Collaborative engagement	Collaborative engagement in K4DD is embedded and iterative, with FCDO staff co-creating evidence through ongoing dialogue and expert sessions. In contrast, the Evidence Fund relies on structured, project-based collaboration, where external researchers engage periodically through formal commissioning and milestone reviews. Therefore, the management arrangements will differ across both programmes.
	Optimised research administration, operational efficiency and technology integration	Given the difference in scale and timelines of the two programmes, the administration will be more agile for K4DD and more structured for EF.

Box 1: Expected differences in the application of the VfM framework between the two programmes		
	Stakeholder & institutional knowledge and continuous learning and improvement	K4DD fosters continuous learning through embedded collaboration, allowing stakeholders to adapt and refine decisions in real time. In contrast, the Evidence Fund emphasises structured learning cycles, where stakeholder insights are captured at key milestones. In K4DD, institutional knowledge is built dynamically through iterative engagement while in EF, institutional memory is shaped through formal documentation and evaluation. In applying the performance standards, there should be a recognition of both, the continuous learning trajectory and institutional knowledge.
Effectiveness	**Instrumental & Conceptual impacts**	EF fund projects may be more likely to result in instrumental impacts, particularly if they involve testing pilot interventions or generating new evidence. K4DD projects may be more likely to generate conceptual impact, by exposing decision-makers to wider bodies of evidence to consider.
	Process Impact-Capacity-building and Connectivity	Both programmes should be able to deliver limited capacity strengthening, but potentially in different forms. EF projects may be more likely to strengthen technical capacity, while K4DD may be more likely to contribute to broader understanding of research and its relevance. EF projects are more likely to generate connectivity, as they are more likely to involve local research teams that may respond to opportunities for further collaboration.
Equity	**Catering to a diverse user base**	The EF is designed to respond to the strategic priorities identified by FCDO. As such, it actively filters and prioritises requests based on their alignment with UK development and diplomatic objectives. While it seeks to ensure wide visibility and accessibility across FCDO, equity of usage (i.e. balanced uptake across departments and regions) is not a primary goal of the programme. In contrast, K4DD is available to FCDO staff on a first-come, first-served basis and does not filter requests based on strategic priorities, enabling broader access. Therefore, K4DD will perform better on than for Evidence Fund on equity of usage.
	Diversity within outputs commissioned	Studies commissioned under EF are aligned with the strategic priorities identified by FCDO, which will somewhat limit the topic areas the programme can cover. However, where relevant, ODA-funded EF studies may incorporate intersectional analyses on equity and inclusion within broader topics. This is less applicable for diplomacy and growth-focused non-ODA studies. K4DD, on the other hand, can produce outputs across a wider range of themes, including those related to equity and inclusion, as it does not screen or refuse requests based on strategic priorities. Consequently, diversity of topic coverage and inclusion of equity considerations are more relevant for K4DD, and it will score higher.

Box 1: Expected differences in the application of the VfM framework between the two programmes		
	Diversity of research partnerships and commissioning:	The Evidence Fund uses a more varied approach to commissioning. EF employs open-market procurement for most studies, which allows for the engagement of a broader pool of research suppliers, including local and regional suppliers in the Global South, while also improving transparency. This can support more localisation, greater diversity in research partnerships, and contextual relevance. However, for certain studies classed as 'official sensitive', EF relies on a narrower set of pre-identified providers and a more targeted procurement process, limiting supplier diversity in these cases. In contrast, K4DD commissions work through its fixed delivery consortium, which restricts the range of institutional partners involved. While this model offers greater speed and consistency, it limits the ability to expand or diversify partnerships.
	Accessibility of research outputs	Under the Evidence Fund, only ODA-funded research outputs are expected to be made publicly available. Non-ODA projects on security, humanitarian, or diplomacy related evidence are sometimes not published due to sensitivity and strategic considerations. Under K4DD, decisions on external publication of outputs rest with the commissioning users or customers, resulting in variability in how accessible K4DD outputs are. Consequently, while the EF's accessibility is guided by clearer publication expectations, K4DD's accessibility depends more heavily on users' discretion, leading to less predictable or more uneven public availability of its outputs.
Cost-effectiveness	**Overall stakeholder value**	The overall value of programmes to stakeholders may be different. EF should provide primary evidence directly relevant to the challenges faced. K4DD provides a broader context of evidence with less specific relevance.
	Uniqueness	EF outputs may score more highly on uniqueness. While K4DD outputs can also be expected to meet real information needs, EF outputs should consist of primary research, providing knowledge that could not come from existing sources.
	Scope, depth and equity	As these criteria relate to the impact of the decisions which are informed by evidence, both programmes could score comparably in these criteria.

8.6 Framework application: data sources and approach

The VfM assessment of both the Evidence Fund and K4DD will need to draw on a mix of quantitative and qualitative data sources, including financial records, delivery timelines, and output quality metrics. To strengthen cross-cutting insights across multiple criteria such as relevance, effectiveness, and efficiency- a targeted survey of evidence users and a set of detailed case studies can serve as pivotal sources. These tools capture user satisfaction and practical impact and help assess collaborative engagement and accessibility. Importantly, both initiatives have existing MEL (Monitoring, Evaluation, and Learning) frameworks, and integrating these VfM data sources with their MEL plans will ensure consistency, reduce duplication, and enhance the utility of findings for both internal learning and external accountability. The Tables below provide initial suggestions on evidence and data sources that could be used to inform a VfM assessment undertaken against the draft framework. **Please note that these are indicative and need to be refined according to the programme and project as the VfM evaluation begins to be undertaken (i.e. moving from**

step 4 to step 5). We suggest having discussions with the PROs and FMs of the programme before application of the framework.

Table 5: Evidence sources for Economy

Sub-criteria	Potential evidence
Admin and governance structure: systems for demand identification, project shaping and management	Administrative and financial oversight: Pipeline and milestones tracker (managed by PwC), programme management dashboards such as DevTracker, FCDO budget and financial sheets, Demand Identification and Governance: Standard Operating Procedures, bidding templates and scoring matrix,
Commissioning process, proportionate inputs, effort and rigour	Bidding process: TORs, Invitation to Tender (ITT) packs, RFPs, evaluation criteria, suppliers database (competitive records, supplier cost comparisons), working arrangements with PwC (contracts, Service Level Agreements for due diligence, risk assessment and contracting)
Operational costs	Budget forecasts, administrative costs, programmatic costs including supplier fees, cost allocation schedule, overheads
Transparent fund oversight	Budget to actual comparisons showing delivery to budget or cost savings, contract terms (milestone based, benchmarks), timely financial reporting
Complementarity	Repeat engagement logs, learning journey reports (K4DD), researcher networks, user KIIs

Table 6: Evidence sources for Efficiency

Sub-criteria	Potential evidence
Strategic Relevance and Structured Prioritisation, right research questions	Supplier Bid evaluation forms, TORs/ITT packs, programme strategy documents, records of referrals or declined bids, alignment assessments with UK/HMG priorities. ToC documents, research gap analysis, scoping studies
Accessibility and Coherence	Research design, methodology notes, quality assurance notes or comments on findings
Quality Outputs	QA assessments and audit reports, timeliness trackers (e.g. Gantt charts, milestone logs), downstream uptake data, procuring services from external peer-reviewers, plagiarism and AI check software
Collaborative Engagement	MoUs or partnership agreements, workshop proceedings or meeting minutes, communication logs (timing, responsiveness), survey of users and research partners.
Optimised Research Administration & Tech Integration	Admin process documentation (SOPs), Procurement timelines and tender records, Platform usage logs (e.g. commissioning portals), System performance dashboards
Stakeholder & Institutional Knowledge and Continuous learning and improvement	Stakeholder mapping reports, capacity assessments or collaborations, internal knowledge repositories or briefing notes, M&E review reports, 'Lessons learned' documents, feedback loops and change logs, action tracker follow-up data

Table 7: Evidence sources for Effectiveness

Sub-criteria	Potential evidence
Instrumental impact	Impact stories or reports from direct users and other relevant decision makers, confirming the value and role of research outputs to instrumental decision making. MEL monitoring and evaluation evidence, Annual Reviews

Sub-criteria	Potential evidence
Conceptual impact	Same as above
Process impact: Capacity building and Connectivity	Assessments of evidence use capacity, evaluation capacity building (ECB), staff trainings, reporting from research users Increased engagement by decision makers in making future requests for evidence or commissioning evidence and synthesis through other routes. Reports from decision makers and researchers that attest to the value of their relationship going forward. Further collaborative projects or engagement between researchers and decision makers initiated independently of the programme.

Table 8: Evidence sources for Equity

Sub-criteria	Potential evidence
Catering to a diverse user base	Research Supplier database, metrics on localisation, PwC Pipeline tracker, data on programme users, disaggregated by FCDO directorate, cadre, region or location of work, ODA/non-ODA, and job role - and analysed over time to examine change in distribution Include questions on awareness and access in user surveys or feedback forms e.g. 'How did you learn about this service?' or 'Were there any barriers you faced when accessing this service?'
Diversity within outputs commissioned	Data on thematic and geographic spread of outputs commissioned - to examine distribution and gaps Review of content to assess if GEDSI issues are incorporated or addressed meaningfully. This can also be collected and/or triangulated via user surveys or feedback forms Review of types of references and sources cited in research outputs (by type of source)
Diversity of research partnerships and commissioning	Analysis of supplier or partner database for contracted projects, with details on location, size of organisation, type of organisation, and contractual role (e.g. lead firm, subcontractor) Detailed feedback from users and/or partners on experiences working with the programme, analysed by supplier type, assessing barriers and capacity-building outcomes
Accessibility of research outputs	Website and/or social media analytics for publicly available (ODA) outputs Data to track variations in the type of outputs requested and/or produced (e.g. from ToRs used for procurement, user feedback) User feedback on outputs, and how easy they were to understand, share, and use (via interviews and open-ended survey questions).

Table 9: Evidence sources for cost-effectiveness

Sub-criteria	Potential evidence
Overall stakeholder value	Stakeholder satisfaction surveys, KIIs, IDIs with appropriate stakeholders
Uniqueness	Gap analysis reports, User surveys and interviews
Scope, depth and equity	Case studies of the potential impact of policies influenced by research outputs, impact libraries, annual reviews Reports from decision makers on the potential scale of impact resulting from conceptual impact of research outputs Impact assessment from business cases or other policy documents.

Sub-criteria	Potential evidence
	Evaluations of programmes informed by research outputs

8.7 Application of the framework to other programmes

Applying the Value for Money (VFM) framework requires a deliberate and strategic approach that strengthens the broader Monitoring, Evaluation, and Learning (MEL) system. To apply the framework, one should begin by defining a clear value proposition that reflects the programme's strategic intent and stakeholder expectations. This should be followed by adapting the VfM criteria to reflect programme-specific attributes, such as delivery mechanisms, geographic scope, thematic focus, and institutional arrangements. Performance standards can then be calibrated to reflect what constitutes 'excellent', 'good', 'adequate', or 'poor' performance in that context, drawing on stakeholder consultations and programme documentation. A dedicated workshop with all stakeholders can help in validating the criteria and performance standards so that everything is clear from the beginning. It can be used as a critical platform to unpack these needs, engage stakeholders in defining what "value" means in context, and co-design data strategies that align with VFM principles.

The framework's emphasis on participatory design, adaptive learning and iterative refinement makes it well-suited for programmes that value learning and adaptive management. Its integration of qualitative and quantitative evidence, along with its recognition of trade-offs and contextual constraints, ensures that VfM assessments remain meaningful and actionable. The framework's data collection recommendations provide a practical roadmap for embedding VfM considerations into routine MEL systems. As FCDO seeks to assess whether interventions are not only efficient but also relevant and impactful, it becomes clear that existing data may be insufficient to capture the full picture. This calls for identifying gaps and creating new data sources - whether through targeted surveys, participatory feedback mechanisms, or real-time monitoring tools.[28]

In sum, the VfM framework offers a robust, flexible, and user-informed approach that can enhance the evaluative capacity of demand-driven programmes across sectors and geographies. Its application can support better resource stewardship, more inclusive decision-making, and stronger accountability for impact making it a valuable tool for institutions committed to evidence-informed governance and development.

28 We would not normally assign weights unless there is a very clear basis to do so, which is not the case with respect to these programmes. It is possible to simply report against the five high level criteria separately. If there is a desire to produce a single overall VfM rating, then (if ratings vary between criteria) it is a matter of judgement about how much importance to give each of the five criteria in the overall assessment.

Annex A: Insights from key informant interviews

Programme Relevance and Strategic Utility

1. Programmes are primarily valued by **FCDO policymakers and programme leads**, with high relevance to decision-making through **demand-responsive evidence and advisory services**.
2. They extend their usefulness to:
    a. **Other HMG departments** (e.g. DSIT, DBT, Home Office)
    b. **International donors**, regional hubs, and in-country partners
    c. **Academic institutions** and contracted agencies generating research outputs
3. Advisory services deliver **strategic insights** and time savings by reducing capacity burdens and informing **senior-level decisions**.
4. Some users also reported the programmes being useful for providing access to **networks** (via contracted agencies or academic institutions) which could be useful for future work and/or providing a pathway for developing their knowledge base in new subject areas (as required by their job roles).

Programme Operations

1. Demand driven programmes support programming, policy, and diplomacy, but demand has largely been driven by research advisors dealing with resource constraints faced by their own teams.
2. Tailored evidence products are delivered based on responsiveness, quality, and timeliness.
3. Annual reviews and feedback loops have prompted a shift toward proactive **commissioning**.

Value for Money (VfM) and Assessment Challenges

1. VfM is framed through the 4Es (economy, efficiency, effectiveness, equity), with sustainability as a cross-cutting lens.
2. While economic value remains difficult to quantify, programmes offer tangible and intangible returns via influence on policy and strategic direction.
3. Cost considerations span:
4. Researcher time, staff overheads, access to journals, Fund Manager fees, graphic design, dissemination travel are the main input costs

Programme Efficiency and Quality Assurance

1. Outsourcing to delivery partners (e.g. PwC, IDS) has improved delivery speed and administrative efficiency to some extent. However, research commissioners do need to provide substantial time to shape and QA the research outputs, in collaboration with the Research Advisers.
2. Users reported providing feedback through survey tools (but Annual Review documents state that response rates for these are low).

Capacity and Commissioning Dynamics

1. Decentralised commissioning reflects diverse user needs across ODA and non-ODA teams.
2. Bid templates and prioritisation windows aim to streamline demand; some flex-funding enables agility.
3. Commissioning is most effective when backed by strategic rationale, avoiding "interest-only" requests.
4. Triaging requests remains a challenge due to limited visibility across overlapping evidence facilities.

Research Themes and Innovation Integration

1. Commissioned work spans diverse areas like organised crime-climate nexus and gender equity gaps.
2. K4DD has been used by some users as a scoping tool to guide follow-up inquiries/requests through EF.
3. Increasing attention to equity and representation in commissioned research -especially from Global South voices - is emerging, and programme team has conducted market engagement exercises to support this.
4. Use of **AI tools in administrative and research processes** is seen as a growing enabler of efficiency.

Coordination and Engagement

1. Collaboration with other FCDO teams is required to ensure complementarity, not duplication.
2. Dissemination plans, proactive publication permissions, and internal "journal clubs" promote learning and sharing within the programme teams
3. Stronger alignment between government and FCDO priorities was cited as key to uptake and relevance.

Annex B: Key Informant Interviews: Details

Stakeholder Group	Serial No.	Name	Role
SRO	1	Helen Dewberry	SRO for the Evidence Fund
Administers of similar programmes	2	Claire Lewis	Helpdesk, Nature Facility, FCDO
	3	Rachel Lineham	Evaluation Unit, FCDO – SRO for Evaluation Quality Assurance and Learning Service
Hub leads (Regional and/or Evaluation Unit) / Research Advisers	4	Jessica Vince	Head of FCDO Evaluation Unit (Analysis Directorate)
	5	Emmeline Skinner	Research and Innovation Adviser, East Africa Research Hub, Evidence Fund
	6	Georgia Watson	Non-ODA research adviser for the Evidence Fund
Researchers/ Academics	7	Kathryn Oliver	Professor of Evidence and Policy, London School of Hygiene and Tropical Medicine, K4DD stakeholder group
	8	Jonathan Breckon	Visiting Senior Research Fellow, University College London
	9	John Lavis	Independent Researcher
PRO	10	Margaret Vasu	Programme Manager, South and Southeast Asia Research & Innovation Hub, Evidence Fund
	11	Ijeoma Agwu	PRO K4DD
FCDO Users	12	Gwilym Jones	East Africa Climate and Environment Adviser, British High Commission, Kigali, EF user
	13	Miranda Alison	Evidence and Evaluation Lead | Conflict and Atrocity Prevention Department [Migration and Conflict Directorate] and EF user
	14	Alan Whaites	Senior Governance Adviser | Centre of Expertise – Governance Platform [Governance Department, Development and Open Societies Directorate] and K4DD user
	15	Sarah McAveety	Serious and Organised Crime Governance Adviser | Transnational Serious and Organised Crime Department [National Security Directorate] and K4DD user
	16	Anjana Seshadri	Research and Innovation Adviser, South and Southeast Asia Research and Innovation Hub and K4DD user

Stakeholder Group	Serial No.	Name	Role
	17	Nankling Danfulani	DBT Trade Adviser, Technology, Education and Creatives Lead, Nigeria and EF user
		Oyinkansola Akintoye-Bello	Country Director, Nigeria and EF user
	18	Chavit Uttamachai	Science and Technology Adviser, Thailand and EF user
Govt Proxy	19	Sam Kumar	Senior Adviser, Economics, Climate and Development, British Deputy High Commission in Chennai and EF user who works closely with the state government of Tamil Nadu

Annex C: Thematic summary of the value proposition of Evidence Fund and K4DD Programmes based on workshops

Value to whom and how?

Critical evidence-based support to primary stakeholders:

1. UK Government Teams (FCDO programme teams, policy teams, Department of Business & Trade, Department of Health and Social Care); informs policy decisions and funding allocations.
    1. Saves capacity by outsourcing research and advisory services, enabling faster decision-making.
    2. Cross-sector learning within FCDO.
    3. Equipping individuals and teams with evidence-based decision-making skills.
    4. Dept Health uses for recruitment policies (e.g., nurse migration)
2. UK businesses and academia: provides market intelligence (e.g., UK Growth Mission) and funds research partnerships.

Other stakeholders:

1. Partner Governments & Regional Actors (e.g., ASEAN): strengthens development planning and diplomatic engagement through contextual evidence.
2. Development Partners: facilitates collaborative policymaking via shared insights.

How do the programmes address inequities?

1. Gender Equality, Disability, and Social Inclusion (GEDSI): embedded in projects like health and climate research, promoting opportunities for women and marginalized groups.
2. Broader or Intersectional Aspects: expanding equity beyond gender and disability to include class, geography, digital accessibility, linguistic and cultural accessibility.
3. Localising Research Partnerships: prioritises regional researchers (e.g., Sarawak, Malaysia) to enhance contextual relevance and build local capacity.
4. Public Accessibility: wider knowledge dissemination beyond privileged institutions.
5. Challenges: limited access to local partners in regions like East Africa and underrepresentation in emerging tech fields.

Outcomes

Attribution is challenging and there can be long time scales between research and policy outcomes such as:

1. Policy Integration: Research directly informs UK Country Business Plans, diplomatic strategies, and partner government policies. Direct (instrumental) policy impact is rare.

2. Conceptual impacts: shifts in thinking and learning culture which may not lead to immediate decisions but are still significant outcomes.
3. Capacity Building: enhanced evidence literacy and analytical skills among FCDO teams and stakeholders.
4. Network Development: strengthened collaborations between policymakers, researchers, and regional actors.
5. Sustainability: long-term shifts in institutional learning cultures and evidence use beyond project timelines (identified as an aspiration or area for improvement).

Ways of working to maximise value

1. Collaborative Engagement: partnering with stakeholders as equal contributors, not just service providers.
2. Structured Prioritization: demand-driven research aligned with UK priorities, structured bidding processes, use of bid templates, refining research questions for actionable insights.
3. Technology Integration: leveraging technology and AI for evidence synthesis.
4. Adaptive Fund Management: Flexible resource allocation and streamlined reporting to reduce administrative burdens.
5. Feedback loops/Continuous Learning Mechanisms: sharing lessons between established and newer hubs; biweekly meetings for monitoring progress; journal clubs.

Stewardship of resources

Resources Invested:

1. Financial: Researcher time, administrative costs (e.g., PwC oversight), and fund maintenance.
2. Human: User co-production time, peer review networks, and supplier partnerships.
3. Intangible: Relationship-building, network capital (e.g., long-term value of trust)

Stewardship Priorities:

1. Cost Efficiency: Balancing rigour with speed, leveraging established networks to reduce costs.
2. Equitable Partnerships: Prioritizing regional researchers and diversifying supplier databases.
3. Transparency: Aligning research with CDEL R&D standards for methodological rigour.

Annex D: Operationalising VfM Assessment Across a Large Portfolio

This framework is intended to provide VfM assessments for the entire programmes. Some of the assessments are based on information that is relevant only at the programme level. However, both programmes have a large number of diverse projects that generate their primary outputs. Some assessment will be done at the project level and aggregated to give programme level judgements. In some cases, data can be collected (and may already exist) for all projects. In other cases, where more detailed or specific information is required, then a sample of projects, or a smaller set of case studies, may be used. Designing all of the details of the data collection approach is beyond the scope of this report and belongs with step 5 of the process, as the implementation of the VfM assessment begins. However, this annex outlines general principles that apply, as well as proposing some core studies that are likely to be useful.

The data collection approaches should be driven by collecting the right evidence for each criterion and sub-criterion i.e. finding the 'best fit' data that can be realistically collected to inform a judgement against the rubric. There may be some iteration with the details of the rubric as this is done. Data required will include both quantitative (e.g., cost and expenditure reports, output indicators) and qualitative (e.g., interviews, surveys, case studies) data. Multiple sources should be used to triangulate findings and provide a comprehensive view.

General principles

1. Use Sampling Where Necessary

- **Representative Sampling:** Rather than assessing all 250 projects individually, select a representative sample based on geography, thematic area, and budget size. This may be stratified by other characteristics such as the type of output produced.

2. Use a Tiered Assessment Approach

- **Portfolio-Level Analysis:** Use aggregated financial and output data to assess VfM across the entire programme.
- **Project-Level Deep Dives:** Conduct detailed VfM assessments for a smaller number of strategically selected projects to illustrate and understand performance and trade-offs. This may be for a representative sample, if resources are available, or for a smaller number of case studies if necessary, and depending in part on the depth of analysis required.

3. Use Existing MEL Infrastructure

- Where possible, data collection should align with and draw on **existing Monitoring, Evaluation, and Learning (MEL) systems** to avoid duplication and make use of existing date and established processes.

- **Leverage Existing Evaluations and Reviews:** Where projects have undergone evaluations or reviews, extract relevant VfM evidence rather than duplicating effort
- **Integrate New Data Requirements with existing MEL Systems:** Where new days is required, wherever possible align VfM data collection with existing reporting cycles and tools to reduce burden. This is particularly important if the framework is likely to be applied multiple times for the two programmes.

4. Standardise Data Templates

- Where these do not already exist, develop simple templates for project teams to report cost, output, and outcome data in a consistent format.
- Include space for qualitative reflections on efficiency, effectiveness, equity, and sustainability.

5. Triangulation and Interpretation

- Combine quantitative data with qualitative insights from stakeholders to understand trade-offs and contextual factors.
- Use case studies to illustrate how VfM manifests differently across project types.

Specific studies and data collection exercises

The details of specific studies and data collection exercises to be undertaken depends in part in what is already available through existing programme MEL systems, which appear to be continuing to develop at the time this report was produced. Nevertheless, some specific studies are likely to be required, in addition to the general document review and cost data analysis mentioned in the main text. They are as follows.

1. **Key Stakeholders Survey**

   Evidence to inform assessment against a number of the VfM criteria requires a **survey of key stakeholders.** This should include particularly the users of the research, who will provide critical feedback on ways of working, perceived quality and relevance and use of the outputs, amongst other things. It should also include research producers, who will provide information on programme commissioning and management processes and on engagement with users, amongst other things. This should probably be undertaken as an online survey using largely closed questions, including Likert scales to record respondents' ratings of various aspects of performance. It should also include some space for qualitative responses. It can therefore be undertaken at scale, potentially using a 'take all' sample of all users and producers who have been involved with the programmes in a specified interval – for example, the preceding 12 or 24 months. It could also use a smaller, stratified sample if a take-all sample were difficult for any reason. It should be structured around individual research projects, so that data can be used at project level as well as at the user level. The details of what is included in this survey will depend on what information is already collected through routine programme MEL data.

2. **Additional In-depth KIIs.**

A large-scale quantitative survey should be complemented with **in-depth interviews** (or, if appropriate, focus groups) to gather more detailed information. This will enable the assessment to better understand the quantitative survey findings and to take into account more nuanced user and producer perspectives on a range of the VfM criteria. It could be undertaken on a purposive sample of key stakeholders, selected according to a set of criteria that will provide a diverse range of experiences and perspectives.

3. Project case studies

A small sample of projects, drawn either randomly with stratification or purposively to ensure a diversity of experience, can be used to inform assessment against a number of criteria – for example, against elements of quality and accessibility – as well as to provide more detailed and nuanced evidence and triangulation for assessment against other criteria.

4. Following Up on Impact

Programme MEL data, the user survey and the KIIs should all collect whatever information they can on the use of research project outputs (elements of effectiveness) and the consequences of that use (an element of cost-effectiveness). However, there are limits to the depth to which this information can be collected in this way, particularly through the first two instruments. The VfM assessment should also consider **compiling specific 'impact stories' / case studies of impact**, which follow up on and describe in-depth potentially important cases of evidence use that are identified through the other channels. This will be valuable in understanding effectiveness and particularly in assessing elements of cost effectiveness.