



# Cost–Benefit Analysis (CBA) or the Highway? An Alternative Road to Investigating the Value for Money of International Development Research

Heidi Peterson<sup>1,2</sup>

Accepted: 4 August 2022 / Published online: 21 October 2022

© European Association of Development Research and Training Institutes (EADI) 2022

## Abstract

Research for development (R4D) funding is increasingly expected to demonstrate value for money (VfM). However, the dominance of positivist approaches to evaluating VfM, such as cost-benefit analysis, do not fully account for the complexity of R4D funds and risk undermining efforts to contribute to transformational development. This paper posits an alternative approach to evaluating VfM, using the UK’s Global Challenges Research Fund and the Newton Fund as case studies. Based on a constructivist approach to valuing outcomes, this approach applies a collaboratively developed rubric-based peer review to a sample of projects. This is more appropriate for the complexity of R4D interventions, particularly when considering uncertain and emergent outcomes over a long timeframe. This approach could be adapted to other complex interventions, demonstrating that our options are not merely “CBA or the highway” and there are indeed alternative routes to evaluating VfM.

**Keywords** Value for money · Research for development · Economic evaluation · Rubrics · Constructivist evaluation

## Résumé

Le financement de la recherche en développement (en anglais : « research for développement », R4D) doit, de plus en plus, prouver son rapport qualité-prix (en anglais : value for money, VfM). Cependant, la domination des approches positivistes envers l’évaluation du VfM, tels que l’analyse coûts-bénéfices, ne prennent pas en compte la complexité des fonds R4D, et risquent l’affaiblissement des efforts de développement transformationnel. Cet étude postule une approche alternative à l’évaluation du VfM, utilisant le Global Challenges Research Fund et le Newton Fund, au Royaume-Uni, comme études de cas. Basée sur un approche constructiviste envers l’évaluation des

---

✉ Heidi Peterson  
heidi.petersonn@gmail.com; ap2702@bath.ac.uk

<sup>1</sup> Clear Horizon, 132B Gwynne St, Cremorne, VIC 3121, Australia

<sup>2</sup> University of Bath, Claverton Down, Bath BA2 7AY, UK



résultats, notre approche applique une critique de pairs, établie sur la base d'une rubrique développée de forme collaborative, a un échantillon de projets. Notre approche est plus adéquate à la complexité des interventions R4D, particulièrement si on considère leurs résultats incertains, qui ne deviennent apparents que sur un laps de temps assez long. Cet approche peut être adaptée à d'autres interventions complexes, démontrant que nos options ne sont pas seulement limitées au CBA, et qu'il y a effectivement d'autres possibilités pour l'évaluation du VfM.

## Introduction

“Foreign aid worth £735 million is squandered” splashed the Daily Mail (Drury 2019, para 1) following the UK government's Independent Commission for Aid Impact's (ICAI) investigation into the Newton Fund, a high-profile research for development (R4D) investment. Sensationalist headlines aside, government public spending is heavily scrutinised. This scrutiny has increasingly focused on whether government programmes are ensuring value for money (VfM), and the Daily Mail was highlighting ICAI's (2019, p. 3) findings that the Newton Fund has “no coherent approach to VfM”. There are many definitions of VfM; of particular relevance for research and development in the UK are the National Audit Office's (2016, p. 1) “optimal use of resources to achieve intended outcomes” and ICAI's (2011, p. 1) “the best use of resources to deliver the desired impact”. Put more simply, are we getting “bang for our buck”? However, being able to ensure VfM necessitates the ability to evaluate it.

Assessing VfM is a form of evaluation, which is related to social research, but is distinct in that it, by definition, answers evaluative questions about the quality and value of something (Davidson 2005; Patton 2002; Scriven 2012). Indeed, evaluation usually requires investigation of “what's so” to determine “so what?”, and often uses social research methodologies (Patton 2002; Davidson 2014). Given its relationship with research, evaluation invites many of the same debates on research philosophies and methodologies, and VfM is no exception. Research philosophy or paradigms comprise ontology (beliefs about the nature of reality) and epistemology (what counts as knowledge about reality). VfM has been dominated by a positivist philosophy, which asserts that there is one objective reality and tends towards methodologies that use empirical observation of facts (Shapiro and Schroeder 2008; Shutt 2015). This has led to the dominance of methods such as cost-benefit analysis (CBA) for assessing VfM. However, this paper will argue that a positivist philosophy, with methods such as CBA, is not appropriate for investigating the VfM of complex government funds. Rather, a constructivist philosophy is more appropriate as it allows stakeholders to explicitly construct what value means to them. This paper will demonstrate this through the case studies of the Newton Fund and the Global Challenges Research Fund (GCRF).

The Newton Fund was announced in 2014 as a £375 million investment over five years to develop research and innovation partnerships to promote the economic and social development of fifteen partner countries, delivered by the UK Department for Business, Energy and Industrial Strategy (BEIS). Then, in 2015, the Newton Fund was increased to £735 million, and GCRF was created: £1.5 billion over five years to



address problems facing developing countries, while developing the UK's ability to deliver cutting-edge research.

However, in November 2020, following the devastating economic impacts of the Covid-19 pandemic, the UK government announced that it was reducing its commitment to the UK aid budget. As a consequence of this, the funding for GCRF and the Newton Fund was reduced by £120 million for 2021–2022 (UKRI 2021), with significant consequences for work funded through GCRF and the Newton Fund. Despite this, the substantial investment in the funds thus far still provides a valuable opportunity to learn about the VfM of complex R4D funds, which can be used to improve future funding. Moreover, in the context of a heavily constrained spending environment, having sound methodologies for assessing the VfM of R4D is increasingly important, not only to be accountable to the UK taxpayer, but for the people in developing countries who are supposed to benefit from these funds.

Therefore, this paper sets out the potential approach that BEIS developed over 2019 and 2020 for evaluating the VfM of GCRF and Newton. BEIS piloted the approach using the Newton Fund in 2020 and 2021, and the approach was then adapted for the external evaluation of GCRF. This paper does not explore the pilot process or the external GCRF evaluation, but rather focuses on the rationale for the approach and how BEIS envisaged it would be used at a fund level (which has not been fully implemented at the time of writing).

This paper is structured as follows. Firstly, I review the literature, focusing on three key issues: contested understandings of VfM, the complex nature of R4D and the limitations of CBA. I then present the alternative approach, which was developed in collaboration with the BEIS GCRF and Newton policy team and wider stakeholders.<sup>1</sup> The approach is essentially a collaboratively developed, rubric-based peer review of a stratified sample of GCRF and Newton projects, the results of which may then be aggregated to appropriate levels. I contend that this approach is more sensitive to the complex nature of R4D and is more able to facilitate learning to improve the VfM of the funds. While developing this approach, a colleague warned that it would be difficult to challenge the norms that posit VfM analysis as “CBA or the highway”. Therefore, I ultimately seek to demonstrate that there are alternative roads, other than relying purely on CBA, to more appropriately evaluate the VfM of R4D and other similarly complex interventions.

## Challenges for Evaluating the Value for Money of Research for Development

Choosing an evaluation design is generally influenced by the three factors: the evaluation questions, evaluation attributes and available designs (Stern 2015). These provide a helpful framing for the challenges identified in the literature for evaluating the

---

<sup>1</sup> I worked on this approach as a UKRI employee, sponsored by BEIS, in 2019 and 2020 and I was further involved in the pilot process in an advisory capacity after leaving BEIS. However, the views expressed in this article are my own and do not necessarily reflect the views of UKRI, BEIS or its partner organisations.



VfM of R4D funds. Firstly, when considering the evaluation question of whether an R4D fund represents VfM, we are faced with nebulous and contested understandings of value. Secondly, R4D funds have several attributes that make investigating VfM challenging. Finally, traditional evaluation designs for VfM are not necessarily the most appropriate in this context.

### **Evaluation Questions: Defining the “Value” in Value for Money**

A key issue in designing this evaluation is the contested definitions of VfM, which is reflective of a wider debate around assessing the value of research. Whereas publicly funded research was historically viewed as implicitly valuable, there has been an increasing focus on demonstrating its explicit value back to society (Bornmann 2013). Research was previously evaluated by its academic excellence but is now also assessed against utilitarian principles, particularly in terms of its contribution to social, economic, cultural or environmental capital, despite a lack of understanding about how to measure this (Ibid.).

In the case of ODA-funded research, there is also the question of value to whom? In domestic UK public spending, the taxpayer and the beneficiaries are broadly the same groups of people, but in ODA, these two groups are separate (Shutt 2015). This becomes more complicated in funds with a dual purpose like GCRF: to benefit people in developing countries but also to improve the capability of the UK to conduct excellent research (BEIS 2016). This creates challenges for evaluation: how much value should be apportioned to developing country recipients versus the UK for the intervention to be considered successful at achieving its dual objectives (Keijzer and Lundsgaarde 2018)? Defining what is meant by value, and value to whom, has significant implications for its evaluation.

### **Attributes of the Evaluand: Recognising Research for Development as Complex**

There are specific attributes of R4D, and research more generally, that create challenges for investigating value and VfM, many of which can be understood through a complexity lens (Douthwaite and Hoffecker 2017). Ramalingam et al. (2008) explain that complexity can be broadly understood as interconnected, interdependent, dynamic, and unpredictable systems, in which change is nonlinear with feedback loops, and where agents continuously react to the system and each other (Ramalingam et al. 2008; Bamberger et al. 2016). R4D interventions take place in such systems; each project is situated in particular socio-economic, political, and cultural circumstances, often alongside and interacting with many other interventions (Ofir et al. 2016). Each project has with multiple stakeholders, often with divergent agendas, who are involved in the production of knowledge, its translation, and impact (Ibid.). Bamberger et al. (2016, p. 5) provide a helpful framework for considering different sources of complexity in an evaluand, including the nature of change and causality, and the characteristics of the intervention itself, which we will briefly consider in relation to R4D.



## The Nature of Change and Causality

The nature of change in R4D creates significant evaluation challenges (Douthwaite and Hoffecker 2017), and one of the critical challenges is the nonlinear and uncertain pathways to achieving impact (Martin 2011). Research activities interact in different ways within complex systems, and often have many feedback loops and unexpected results (Roling 2011). In particular, there can be specific outcomes that only emerge during implementation. Emergent outcomes may not be predictable in advance, which makes identifying them difficult, and precludes some pre- and post-comparisons (Ibid.). The uncertain nature of R4D also means that some projects are unlikely to achieve their intended outcomes (Hollow 2013). Supporting transformational research means accepting a degree of risk and an appropriate level of failure, given that some research will be highly successful (Ibid.). Projects that fail to achieve specific outcomes can still have value in contributing to a body of knowledge and allowing other researchers to build on their lessons. Large research funds like GCRF and Newton use a portfolio approach (funding many projects with the expectation that some may not be successful) to manage this risk (Hollow 2013; IDRC 2017) and the evaluation needs to account for this.

Furthermore, the long timeframe for the full benefits of research to be realised is a significant challenge, as the full impact of research cannot be assessed until many years after the project (Ekboir 2003; Bornmann 2013). It is important to evaluate VfM during implementation of funds like GCRF and Newton in order to learn and improve. However, the literature advises doing so with caution, as one may find skewed outcomes that ignore the longer-term and potentially more significant impacts (Bozeman and Youtie 2017). This may then distort future projects if researchers feel pressure to emphasise types of research that will demonstrate short-term results (Martin and Tang 2007; Stern 2016).

Even where impacts can be identified within a specific timeframe, it is difficult to attribute them to any one R4D intervention. As Martin and Tang (2007, p. 3) explain, research projects do not operate in silos, and they often build on each other. This creates a chain of escalating uncertainties when evaluators try to trace a particular benefit back to a specific research project. The diffusion of benefits further complicates this; with research outputs increasingly accessible online, innovation has become more globalised. Innovation in one country is likely to draw upon outputs of research in other countries; therefore, investing in research in one country is likely to impact many other countries (Ibid.).

## The Nature of the Intervention

The complexity detailed above is for R4D in general, but it is amplified when we consider the nature of GCRF and Newton as interventions. “Complicated” variables of the intervention can increase its complexity, including: organisational delivery, scale and geographic spread, diversity of the target population, number and diversity of programme components, and the range of programme objectives (Rogers 2008). When considering all of these variables, GCRF and Newton are indeed complicated



funds. GCRF and Newton are funded by BEIS but delivered through seventeen different Delivery Partners, such as UK Research and Innovation (UKRI) (including its constituent Research Councils), the Meteorological Office and the UK Space Agency. GCRF alone has over seventy programmes (each with different objectives ranging from sustainable livelihoods, through refugee crises, to gender inequality) and over 2000 individual R4D projects operating in over 77 countries with thousands of partner organisations.

Despite this greater understanding of complexity, Douthwaite and Hoffecker (2017) contend that many evaluations, particularly by major international funders of R4D, remain complexity blind. In VfM assessments, this is partly due to the unsuitability of dominant methods to account for this complexity, to which we will now turn.

### Available Designs: The Suitability of Traditional Econometric Methods

While there have been efforts in recent years to push the research evaluation field forwards to recognise complexity (Ibid.), such efforts tend to focus on impact evaluation rather than explicit VfM assessment. In the field of VfM, CBA, and associated methodologies, are still seen as the gold standard (King 2019a)—although it should be noted that funding decisions do not tend to be made on CBA alone (Peterson & Skolits 2020). CBA essentially tallies benefits and costs (all expressed in units of money in present value terms) to calculate the net benefit. However, the limitations of CBA have long been recognised; critiques fall into three categories: philosophical, epistemological and methodological.

Philosophically, there is the choice to use the weighing of economic costs and benefits to make public policy decisions. This foundation in bounded utilitarianism divides opinion from the get-go (Frank 2000; Sen 2000). Then, CBA is premised on normative choices and assumptions drawn from economic theories. CBA generally holds that individuals are rational, so preference satisfaction leads to individual welfare and, therefore, the aggregation of individual market choices is an appropriate investigation of societal welfare (Kopp et al. 1997; Næss 2006; Hausman and McPherson 2008). The validity and ethics of treating individuals as rational consumers, rather than citizens with rights or moral codes, is debated (Nussbaum 2000; Sen 2000; Ackerman and Heinzerling 2002).

Epistemologically, advocates of CBA contend that its advantage is that it is rigorous, transparent and objective, as ideally all costs and benefits are tallied against each other (Kopp et al. 1997; Ackerman and Heinzerling 2002). However, this premise is flawed given the subjective decisions that are made when applying this approach (Ackerman and Heinzerling 2002; Næss 2006; King 2019b). Decisions are made in selecting the timeframe of the benefits, deciding which costs and benefits to include, choosing which discount rate to use, and which methodologies to use for intangible benefits (Næss 2006; King 2019b). Moreover, CBA is a technical endeavour that is not particularly transparent to stakeholders who cannot see or debate the underlying decisions, and it also precludes stakeholders from actively participating in the approach (Ackerman and Heinzerling 2002; Næss 2006).



Methodologically, there is a robust critique of CBA in the literature. I have selected four critiques that are pertinent for R4D projects. Firstly, CBA is generally used ex-ante or ex-post as it requires all costs and benefits to be able to be predicted or known (Sen 2000). This is difficult for R4D projects; as already noted, its unexpected and emergent nature makes it challenging to predict all benefits ante-post. The impacts of research often take a long time to be realised, meaning that an ex-post CBA may not capture them and may lead to false conclusions. Moreover, it is difficult to attribute a particular benefit wholly to one research project; as discussed, projects tend to interrelate. A VfM evaluation would need to be conducted during implementation if one wants to be able to improve during the intervention, to which the use of CBA is not conducive.

Secondly, CBA struggles to incorporate intangible and non-quantifiable benefits (Kopp et al. 1997; Ackerman and Heinzerling 2002), of which there are many in research (Martin and Tang 2007). While methods have developed to try to incorporate them, the validity of attempts to quantify and place a monetary value on these benefits is contested (Nussbaum 2000; Frank 2000; Ackerman and Heinzerling 2002; King 2019b). The most commonly used technique is assessing the willingness-to-pay of individuals. However, studies have shown that using the willingness-to-pay measure differs within the same group of people depending on when they are asked (King 2019b). Moreover, the willingness-to-pay measure estimates individual willingness, not collective willingness (Næss 2006). Therefore, CBA struggles to quantify the benefit of investing in global public goods (Ackerman and Heinzerling 2002; Sen 2000), which is problematic for a fund like GCRF, which aims to address global challenges such as climate change.

Thirdly, CBA struggles to take equity into account. This is partly philosophical, given that not all decisions that are the most cost-effective are the most equitable. In funds like GCRF and Newton, the concern is that a VfM assessment could disadvantage important research taking place in challenging contexts. The value that one might expect to see for a certain investment may be different; for example, in a fragile or conflict-affected setting versus a more secure environment. It is also in part methodological, as CBA tends to use individual measures of welfare to reach the aggregate social benefit and traditionally no person's welfare would be weighted (Kopp et al. 1997; Ackerman and Heinzerling 2002; Næss 2006). Therefore, CBA does not tend to consider the distribution effects of an intervention, which could unequally benefit a certain group of people. Efforts to weight people's welfare are ethically fraught and require subjective decisions. This is particularly an issue for R4D, where equitable development is a widely agreed principle (Shutt 2015).

Finally, CBA only looks at the benefits or outcomes, not the process by which they were achieved (Hausman and McPherson 2008). There is a crucial VfM consideration in the process and principles of conducting R4D. Consider two research projects that achieve the same outcome. Project A costs more because it uses resources in a way that furthers equitable partnerships, whereas project B costs less but furthers unequal relationships. Project B would give a more favourable CBA, but project A is actually more desirable for GCRF's principle of equitable partnerships (Rethinking Research Collaborative 2018).



In summary, the literature shows that CBA struggles with many R4D projects (Martin and Tang 2007; Bornmann 2013; Bozeman and Youtie 2017). We need an alternative design but as Anderson (2008, p. 1) notes, “few attempts have been made to bridge the gap between methods of economic evaluation and the broader methodological debates about the definition and evaluation of complex interventions”.

### Key Evaluation Questions

While a VfM evaluation implies an inherent question (Is this value for money?), a more nuanced evaluation that recognises the aforementioned issues has the following Key Evaluation Questions (KEQs):

1. What do GCRF and Newton stakeholders define as value for money for the funds?
2. To what extent are GCRF and Newton projects positioned to deliver value for money as defined by stakeholders?
3. What can GCRF and Newton stakeholders learn to improve value for money?

The next section will outline the proposed approach to answering these questions.

### The Approach

The approach uses a collaboratively developed rubric-based review of a sample of GCRF and Newton projects, the findings of which can then be synthesised to appropriate levels. This approach draws on Canada’s International Development Research Centre’s (2017) Research Quality Plus framework (which uses a rubric to assess the quality of different R4D projects) and King and Oxford Policy Management’s (2018) approach (which uses a rubric to assess the VfM in international development programmes)<sup>2</sup>. The steps of the approach are summarised in Fig. 1.

Before delving into the detail of the steps, it is worth considering the approach’s underlying research philosophy. In considering the philosophical home for this approach, it is first important to consider the nature of evaluation. As discussed in the introduction, evaluation differs from research in that it answers evaluative questions about the quality and value of something (Davidson 2005; Scriven 2012). Therefore, there is a philosophical question about determining the “something” (the what’s so?) and a separate philosophical question about determining its value (the so what?) (although both philosophies could be the same).

---

<sup>2</sup> This methodology is related to multi-criteria analysis (MCA) approaches, which similarly determine criteria and require judgement to assess alternatives using the criteria, thereby not requiring monetary valuation of outcomes (Department of Local Communities and Government 2009). While MCA is primarily used prospectively as a decision-making tool to determine the most preferred option from a set of alternatives, Vardakoulis (2013) argues that it could be used retrospectively for VFM evaluation, although it has been rarely used for such. However, MCA tends to use relative weighting methodologies; King (2019b) argues that using an evaluation rubric with ordinal weights (such as is used in this paper) is more appropriate for assessing VFM where there is no clear basis for setting relative weights.





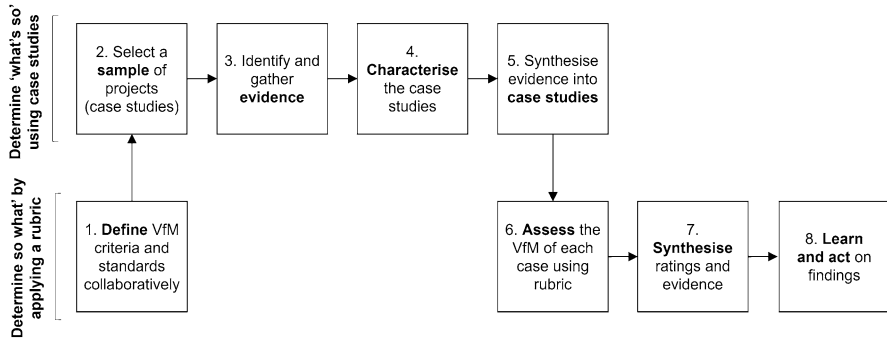


Fig. 1 Overview of the approach (Adapted from King and OPM (2018) and IDRC (2017))

Given the complex nature of R4D, particularly in funds like GCRF and Newton with thousands of diverse projects, different outcomes could be investigated in different ways. Perhaps the outcome could lend itself to a positivist paradigm [where there can be a single, objective reality that can be empirically observed (Shutt 2015)], such as a medical trial. Perhaps other cases, such as a capacity-building network, would be more suited to an interpretivist paradigm [where there can be multiple realities that are socially constructed (Walker and Dewar 2000)]. In the case of GCRF and Newton, most projects are given autonomy in deciding how to generate evidence of their outcomes. This paper is more concerned with the ontology and epistemology of the valuing of these outcomes.

Traditional approaches to VfM such as CBA take a positivist approach to determining value, positing that facts can be empirically determined separately from values (Shapiro and Schroeder 2008). While a particular result of an R4D intervention can potentially be empirically observed, the literature review illustrated that the value of this result cannot be thus observed. Value is not an abstract concept, but rather implies value to somebody or against certain criteria developed by people; the same outcome could be valued differently by different people or criteria (Shutt 2015). For this reason, the VfM approach is probably most at home in a constructivist paradigm, where the value of R4D process and outcomes is constructed by stakeholders. Therefore, Fig. 2 shows the steps of the approach (slightly re-ordered) with the research philosophies.

Constructivist evaluation, pioneered by Guba and Lincoln (1989), has several elements that make it suitable for this approach. It recognises that a plurality of values shape constructions (van de Kerkhof et al. 2010) and acknowledges that values can change as goals shift, knowledge develops, and experience is gained (Walker and Dewar 2000). It also recognises that constructions are inextricably intertwined with context, both the contexts to which they refer and in which they are formed (van de Kerkhof et al. 2010). This is particularly important for complexity-aware evaluation, where contextualisation is key to understanding complex processes and outcomes (Verweij and Gerrits 2013).



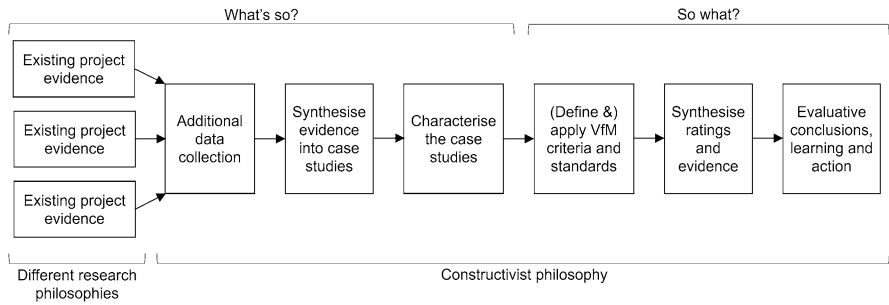


Fig. 2 The approach with underlying research philosophies

Constructivism is similar to interpretivism in that it holds that there can be multiple socially constructed realities but differs in the understanding of the relationship between the evaluator and evaluated. Whereas in interpretivism, the evaluator could be quite separate from the evaluated (while still valuing their subjective interpretations), in constructivism, the evaluator is fully involved in the co-production of knowledge as a form of collaborative inquiry (Walker and Dewar 2000; Abma 2004). Co-production is appropriate for this case because BEIS is responsible for policies that could affect VfM, and the evaluation team sits within BEIS. Constructivist evaluation is oriented towards action (Walker and Dewar 2000; Abma 2004), which is important for this proposal where BEIS aims to facilitate learning with its R4D community to improve VfM.

Given the above, constructivism lends itself to participatory approaches to strengthen the collaborative inquiry. Participatory approaches involve stakeholders, particularly those within a programme, in specific parts of the evaluation. The use of participatory approaches strengthens the ability to answer the evaluation questions; for example, understanding how stakeholders define the value of the funds. It is also important for the wider purpose of the evaluation: to facilitate learning and action. Where BEIS is the funder, it is to be expected that stakeholders may feel concerned about the implications of such an assessment. However, if stakeholders are involved throughout the process, from designing the approach, through defining value, to making sense of findings, the evaluation is better positioned to facilitate a learning environment (Peterson and Skolits 2020).

The choice to design a participatory evaluation is not only for practical benefits, but for ethical reasons. As Barnett and Camfield (2016) explain, evaluation differs from social research in that, if the purpose of the evaluation is to inform decision-making, stakeholders may be directly affected by these decisions. Therefore, stakeholders may have a more direct interest in the results of the evaluation and should be able to engage in a process that will directly affect them. Part of this necessitates an ethical understanding of the ways in which the evaluation interacts with and creates power dynamics (Ibid.). There is an inherent power imbalance in GCRF and Newton where BEIS is the funder, which is likely to be exacerbated here, where BEIS is also the evaluator (or commissioner of the evaluation). It is hoped that a collaborative approach, though not able to completely negate the power imbalance,



**Fig. 3** Rubrics normally have a tabular structure of criteria and standards

Standards	Criterion 1	Criterion 2	Criterion 3
Excellent			
Good			
Adequate			
Poor			

could empower stakeholders to have more influence over the process by which evaluative conclusions are made about VfM and the subsequent decisions. While there are some practical constraints to full participation of all stakeholders (particularly those in low- and middle-income countries (LMIC)), the approach outlined below draws on participatory methods where possible.

### Detailing the Approach

The approach uses two methodologies: case studies (to determine ‘what’s so’) and a rubric (to determine ‘so what?’). Multiple case studies is a methodology in which in-depth information about specific cases is collected, analysed and presented in a narrative format, enabling comparison between cases in a systematic way (Ghauri 2004; Marrelli 2007). Yin (2018) explains that case studies are useful when “how” or “why” questions are asked, when there is little control over external variables and when the focus is on the case in its context – all of which apply here. The advantage of case studies is that they can “offer rich perspectives insights that can lead to an in-depth understanding of variables, issues, and problems” (Marrelli 2007, p.40). This is appropriate for evaluating complexity, as it is sensitive to the diversity, uniqueness and contextualisation of each case (Verweij and Gerrits 2013) as well as the uncertainty, unpredictability and the unexpected (Anderson et al. 2005). Contextualisation is crucial to complexity-aware evaluation (Verweij and Gerrits 2013) and as already noted, it is also important that any VfM assessment does not unduly disadvantage research projects taking place in difficult contexts.

The case studies form the evidence for making evaluative judgements about VfM.

Rubrics (Fig. 3) provide a clear and transparent method for making evaluative judgements through explicit valuation (Davidson 2014; King et al. 2013; Martens 2018). A rubric is essentially a tool that provides an evaluative description of what performance or quality looks like at different, defined levels, which can then be used to convert evidence into explicit conclusions about value (Davidson 2005). The steps below detail the use of these methods in the approach.

#### *Step 1: Define the VfM rubric collaboratively with stakeholders*

King and OPM (2018) recommend developing and validating the criteria and standards in a collaborative manner with stakeholders. This allows multiple stakeholders to reach a shared understanding of what VfM means for them and their intervention and in this case, answering KEQ1 (What do GCRF and Newton stakeholders define as value for money for the funds?). The key stakeholders involved



in this approach were BEIS, Delivery Partners, and a selection of GCRF and Newton researchers and their partners. The process involved brainstorming and selecting the VfM criteria, developing descriptors for each standard of the criterion, and then testing on a selection of projects to refine the rubric in an iterative fashion. The final rubric criteria for the Newton Fund pilot were relevance, equitable partnerships, capacity strengthening, progress on outputs, the likelihood of achieving project objectives, and the likelihood of contributing to fund-level impact—with then a final criterion for the worthwhileness of investment in the light of the other criteria (the full rubric is available in the appendix).

The advantage of developing specific criteria and standards is that BEIS was able to include components of particular importance to the funds. For example, many GCRF and Newton projects involve partnerships between UK and LMIC researchers and research institutions. Facilitating and strengthening equitable partnerships is critical to the funds and has implications for project-level VfM (for example, investing in data storage software that all partners can access). Therefore, equitable partnerships were included as a criterion. The rubric was also developed to be able to account for some of the traditional challenges in assessing research, such as the long timeframe for impact. For example, considering the project's positioning for use of the research (such as targeting potential users) can facilitate a judgement as to whether the project is likely to deliver impact over a longer timeframe than the assessment, rather than judging only on outcomes that have been achieved at the point in time of the assessment.

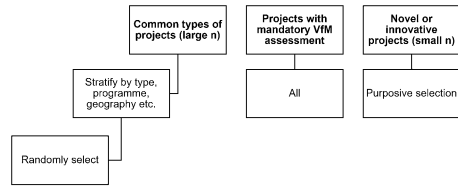
### *Step 2: Select a sample of projects*

Once the VfM criteria and standards have been clearly defined in the rubric, projects then need to be selected as case studies. As GCRF and Newton, respectively, contain thousands of projects, it would be too resource-intensive (nor VfM!) to investigate each project. Case studies tend to use a sampling strategy that aims to generate insights into key issues rather than empirical generalisation from a sample to a population (Patton 2015). Case studies generally use a purposive sample, partly because the sample size is often small (Marrelli 2007). However, GCRF and Newton have a large population size of thousands of projects. From a learning perspective, both typical cases and unusual cases can be useful (Marrelli 2007; Crowe et al. 2011). However, given the complicated nature of GCRF and Newton, there is not yet understanding of what constitutes a typical case in the funds (even within common types of projects).

Therefore, the sampling strategy would have three strands (Fig. 4). The majority of GCRF and Newton projects, comprising of common types of research projects (fellowships, research grants and so on), would be stratified to ensure there is representation across areas of interest, and then randomly sampled to ensure that the best examples are not cherry-picked nor was there a flawed understanding of a typical case. All projects that require a mandatory VfM assessment (generally projects over £5 million) would be included as the second strand. Small but novel types of projects that would strengthen learning if they were included would form the third strand.



Fig. 4 Sampling frame



### Step 3: Identify and gather required evidence

As discussed earlier, most GCRF and Newton projects are given autonomy in deciding how to generate evidence of their outcomes. This generates enormous diversity in the types of data. The advantage of a rubric is that it allows different types of data to become comparable through the same process. For most projects, data collection would involve synthesising existing data (such as project objectives, monitoring and results data and expenditure information) and then triangulating with a semi-structured interview of the project lead (and research team or research partners where appropriate).<sup>3</sup>

Semi-structured interviews allow specific lines of inquiry to be pursued consistently across cases on the criteria set in the rubric. This allows information to be collected on the context of the project, the reasons why projects made decisions relevant to VfM, and the challenges the project faced in trying to ensure VfM. Semi-structuring also allows interviewees to direct certain lines of inquiry that they feel are relevant to VfM, adding a more constructivist approach to the data collection. It also allows the interviewee to identify the unexpected, an essential part of complexity-aware evaluation (Anderson et al. 2005).

### Step 4: Synthesise project data into case studies

The case study data would be used for two purposes: as individual case studies for the assessment of each project's VfM in step 6 and as collective case studies to inform common conclusions in step 7. The analysis of the data would follow two stages, as recommended by Ghauri (2004). For the individual analysis, Ghauri (2004, p. 34) explains that is "hard to describe and explain something satisfactorily unless you understand what this 'something' is". Therefore, the first stage is storytelling to describe the case study's situation and progress. The second stage involves thematic coding for both the individual and collective analysis. Coding involves breaking down the data into specific categories or themes (in this case, using the criteria in the rubric), which can then be analysed in these categories, to achieve a description or understanding of the case study from multiple sources of evidence (Rowley 2002). The thematic codes can then be used with the associated evidence to construct a detailed narrative summary for each case (Iwakabe and Gazzola 2009). The thematic codes would be used again in step 7 to find common conclusions.

<sup>3</sup> Some R4D projects may be suited to a CBA or similar methodology. Where appropriate, a CBA can be conducted at the project level and synthesised as part of the VfM evidence.



**Table 1** Fictional example of project ratings

VfM criterion	Project						Average
	A	B	C	D	E	F	
Efficiency	5	2	1	5	5	4	3.7
Equitable partnerships	1	1	2	5	4	5	3
Likelihood of impact	2	3	5	1	4	4	3.2
Total	8	6	8	11	13	13	9.8

*Step 5: Characterise the projects*

The narrative case summaries prepared in step 4 would enable the characterisation of each project, which serves two purposes. Firstly, it enables the most appropriate application of the rubric to each project (IDRC 2017). Not all rubric criteria may be relevant for every project; characterising awards prior to assessment allows for a more appropriate and nuanced assessment. Secondly, when synthesising the evidence and assessment results, characterisation of the projects allows for a greater investigation and understanding of how these characteristics interface with VfM findings. For example, projects with a high focus on capacity-building can be compared with projects with a low focus.

*Step 6: Assess the VfM of each project using a rubric and peer/expert review*

In the research community, peer review is an important approach for assessing quality, where it is held that experts from the relevant fields of research are best placed to judge the quality or value of research (Ofir et al. 2016; Bozeman and Youtie 2017; IDRC 2017). Peer review is not without critique; however, GCRF and Newton are embedded in the wider system of UK research assessment, where there are significant concerns that assessment allows for subversive control and influence by government over research (Martin 2011). Peer review is viewed as an essential mechanism to limit this (Ibid.). Further, the GCRF and Newton community are highly familiar with peer review, given such review's role in funding applications and the mid-term assessments of large projects. Therefore, BEIS would use panels formed of peer reviewers with additional relevant stakeholders. The panels would use the rubric against the information in the narrative case summaries to assess each project, resulting in a VfM rating for each project.

*Step 7: Synthesise the ratings and evidence to levels of interest*

Synthesis is “the process of combining a set of ratings or performances on several components or dimensions into an overall rating” (Scriven 1991, p. 342). In order to partly answer KEQ2 (to what extent are GCRF and Newton projects positioned to deliver value for money as defined by stakeholders?), the project ratings achieved in step 6 can be synthesised using numerical weight and sum (Davidson 2005). The ratings can be synthesised to levels of interest dependent on the sampling stratification (such as programme, country, type of project), as shown in Table 1.

However, as IDRC (2017) notes, care needs to be taken when aggregating project ratings to different levels to ensure that overall ratings are underpinned by strong



qualitative narratives. The ratings enable BEIS and Delivery Partners to track VfM performance at various levels within the funds, but the narratives allow learning from common themes in the qualitative narratives. Using the characteristics identified in step 4, we can see if these correlate with VfM ratings. For example, in Table 1, if projects A to C had a low focus on capacity-building and projects D to F had a high focus, we could compare any differences and interrogate the qualitative data for potential reasons for this. Using the thematic coding in step 4, other insights can be synthesised across case studies. This allows for BEIS, Delivery Partners and award holders to understand in a more nuanced way what enables and hinders VfM. For example, there may be a common challenge that affects projects' ability to improve VfM, such as a limited time to develop effective partnerships between UK and LMIC research institutions, while projects with good VfM may have leveraged pre-existing relationships to develop strong and equitable partnerships.

There would then be a sense-making workshop with participants to collaboratively investigate findings in order to draw evaluative conclusions. This is an important part of the constructivist approach to valuing, where final conclusions are the result of a process in which different stakeholders participate (Davidson 2005).

#### *Step 8: Facilitate learning and action of findings*

It was important to BEIS and Delivery Partners that this evaluation should lead to learning and action against the findings. An important consideration for this assessment is the portfolio nature of R4D in terms of risk and failure. This is articulated well by Hollow's (2013, p. 25):

“there may be an opportunity [for] a VfM assessment process which is primarily based on portfolio rather than individual programmes, which embraces risk and recognises that failure is a vital part of the learning process”.

Care needs to be taken when feeding back results to individual projects. BEIS and Delivery Partners are unlikely to want to incentivise perverse behaviour in projects merely to score well in a tokenistic fashion, as has been seen in some other research assessment exercises (Martin 2011; Hollow 2013). Rather, BEIS and Delivery Partners want to facilitate genuine, meaningful learning around the VfM of R4D. Therefore, projects would receive the synthesised learnings at the portfolio/fund level rather than receiving individual scores.<sup>4</sup>

Importantly, BEIS and Delivery Partners would be able to adapt based on the findings—in the hypothetical example, BEIS and Delivery Partners could ensure that future projects have sufficient time to build effective partnerships. These sorts of learnings would likely not be evident if only conducting a CBA, where there are usually no qualitative insights to understand why certain projects offer greater VfM than others.

<sup>4</sup> Except in the cases where there are very concerning findings, in which case, a transparent escalation procedure would be followed.



## Critical Reflection of the Approach

The critical reflection of the approach uses Patton's (2002) criteria that evaluations should be meaningful, credible and useful.

### Is the Evaluation Meaningful?

While many aspects of the evaluation contribute to its meaningfulness, of note here is Lincoln and Guba's (1986) concept of fairness. This is the extent to which the constructions, values and opinions of people are genuinely involved in the evaluation. A limitation of this approach is the lack of meaningful participation from beneficiaries. While this approach include some participation from LMIC researchers and partners, it does not involve participation from research beneficiaries. This is for practical reasons (cost and logistics) but given that the funds are ultimately supposed to benefit these people, it is a flaw that they will not be involved in the valuing process.

### Is the Evaluation Credible?

Lincoln and Guba (1986) define credibility as the match between the realities and views of participants and the evaluator's representation of them. A number of strategies identified by Noble and Smith (2015), Crowe et al. (2011) and Abma (2004) have been employed in this approach to increase the credibility, including the use of "thick description" (using rich and verbatim extracts from participants) and data triangulation (where different data and perspectives can corroborate and produce more comprehensive findings). Triangulation is particularly important here, where existing project data are self-reported and may be unreliable.

Perhaps of greater interest here is the credibility of the valuing process, which can be perceived to (and can in fact) be very subjective. Of key concern here is the potential for inappropriate influence of personal preferences and biases (Davidson 2005). This approach has employed several strategies to minimise this, including using explicit valuation (through the rubric) resulting in a transparent application of values that were collaboratively informed by knowledge and experience. This approach also uses informed stakeholders (by providing information through the narrative case summaries, as well as tacit knowledge through project and sector experience). Scriven (1991) notes that rubrics can be affected by shared bias but King et al. (2013) assert that this can be reduced by triangulation (in this case, bringing various groups together on the panel to verify conclusions from different perspectives). Despite this, valuing is an inherently human process and will always be subjective to an extent. However, this approach, like the IDRCs (McLean and Sen 2019), embraces this subjectivity, within the parameters outlined above, to enable stakeholders to make an assessment meaningful to them.





## Is the Evaluation Useful?

Patton (2002) argues that usefulness to direct stakeholders should be the primary criterion for critiquing an evaluation. This paper detailed how the use of participatory methods positions for this, but there are still some limitations. This approach only considers VfM at the project level and does not consider the VfM of other important aspects such as programme delivery and strategic decision-making. These limitations may make the findings less useful for fund-level decision-making, and this approach could potentially be strengthened by a more holistic view of VfM. There is also usefulness in a broader sense. Lincoln and Guba (1986) use the construct “applicability” (instead of generalisability) to consider whether the findings could be applied to other interventions or contexts. Care needs to be taken in reducing complexity to achieve applicability (Verweij and Gerrits 2013) but R4D projects are not unique to GCRF and Newton. It may be that certain VfM enablers or challenges facing GCRF and Newton projects are applicable to other R4D interventions. The approach would provide detailed valuing and contextual information so that others can decide if the findings are appropriate and applicable to their situation (Shenton 2004).

## Conclusion

As large funds, GCRF and Newton present a unique opportunity to investigate what hinders and enables VfM in a wide range of cases. Particularly following the financial impacts of COVID-19, it is crucial to have appropriate methodologies for evaluating VfM. However, policy evaluation, and indeed policy research more broadly, tends to rely on dominant philosophies and associated methodologies, despite the fact that they may not be the most appropriate for every circumstance. In the case of VfM, CBA, underpinned by a positivist philosophy, has been the commanding force for nigh on fifty years. Yet this paper has demonstrated that a constructivist philosophy can be more appropriate for complex funds. This paper developed a novel approach to evaluating the VfM of complex R4D interventions, based on a constructivist approach to valuing, using a rubric and peer review.

The approach developed in this paper is premised on three key tenets. (1) The complexity of R4D projects means that they are inextricably linked to context, and pathways to impact are uncertain, unpredictable and emergent. (2) Using utilitarian methods such as CBA to determine the VfM of such interventions does not appropriately account for this and may risk R4D interventions focusing on short-term, demonstrable and quantifiable results, potentially limiting the transformational potential of funds such as GCRF and Newton. (3) The “value” in VfM is inherently subjective—this is not a limitation, but rather represents an opportunity for stakeholders to collectively construct what value means to them, increasing the meaningfulness of the evaluation. Undoubtedly, these tenets will underpin other complex funds and programmes, not only those found in R4D, and this approach could be adapted to other interventions. Our options are not merely “CBA or the highway”—there are indeed alternative routes to evaluating VfM.



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1057/s41287-022-00565-7>.

## Declarations

**Conflict of interest** I am reporting that I was previously an employee of UK Research and Innovation (UKRI), sponsored by BEIS; the work reported in this article was conducted while I was an employee of UKRI. I am no longer a UKRI employee and do not have a financial interest in publishing this paper.

## References

- Abma, T. 2004. Responsive evaluation: The meaning and special contribution to public administration. *Public Administration* 82 (4): 993–1012.
- Ackerman, F., and L. Heinzerling. 2002. Pricing the priceless: Cost-benefit analysis of environmental protection. *University of Pennsylvania Law Review* 150 (5): 1553–1584.
- Anderson, R. 2008. New MRC guidance on evaluating complex interventions. *British Medical Journal*. <https://doi.org/10.1136/bmj.a1937>.
- Anderson, R.A., B.F. Crabtree, D.J. Steele, and R.R. McDaniel. 2005. Case study research: The view from complexity science. *Qualitative Health Research* 15 (5): 669–685.
- Bamberger, M., J. Vaessen, and E. Raimondo. 2016. *Dealing with complexity in development evaluation: A practical approach*. Thousand Oaks, CA.: Sage.
- Barnett, C., and L. Camfield. 2016. Ethics in evaluation. *Journal of Development Effectiveness* 8 (4): 528–534.
- Bornmann, L. 2013. What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology* 64 (2): 217–233.
- Bozeman, B., and J. Youtie. 2017. Socio-economic impacts and public value of government-funded research: Lessons from four US National Science Foundation initiatives. *Research Policy* 46 (8): 1387–1398.
- Crowe, S., K. Cresswell, A. Robertson, G. Huby, A. Avery, and A. Sheikh. 2011. The case study approach. *BMC Medical Research Methodology* 11: 100. <https://doi.org/10.1186/1471-2288-11-100>.
- Davidson, E. 2005. *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA.: Sage.
- Davidson, E.J. 2014. Evaluative reasoning. Methodological briefs: Impact evaluation 5. Florence: UNICEF Office of Research. [https://www.unicef-irc.org/publications/pdf/brief\\_4\\_evaluativereasoning\\_eng.pdf](https://www.unicef-irc.org/publications/pdf/brief_4_evaluativereasoning_eng.pdf). Accessed 20 Feb 2020.
- Department for Business, Innovation and Skills. 2016. The allocation of science and research funding 2016/17 to 2019/20. London: Department for Business, Energy and Industrial Strategy. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/505308/bis-16-160-allocation-science-research-funding-2016-17-2019-20.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/505308/bis-16-160-allocation-science-research-funding-2016-17-2019-20.pdf). Accessed 30 Jan 2020.
- Department for Communities and Local Government. 2009 Multi-criteria analysis: A manual. London: Department for Communities and Local Government. [http://eprints.lse.ac.uk/12761/1/Multi-criteria\\_Analysis.pdf](http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf). Accessed 2 May 2022.
- Douthwaite, B., and E. Hoffecker. 2017. Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems* 155: 88–102.
- Drury, I. 2019. Foreign aid worth £735 million is squandered on funding studies into jazz and Roman statues, damning watchdog report reveals. The Daily Mail, 8 June 2019. <https://www.dailymail.co.uk/news/article-7118131/Foreign-aid-worth-735MILLION-squandered-funding-studies-jazz-Roman-statues.html>. Accessed: 20 Feb 2020.
- Ekboir, J. 2003. Why impact analysis should not be used for research evaluation and what the alternatives are. *Agricultural Systems* 78: 166–184.
- European Initiative for Agricultural Research for Development. 2003. Impact assessment and evaluation in agricultural research for development. *Agricultural Systems* 78 (2): 329–336.



- Frank, R.H. 2000. Why is cost-benefit analysis so controversial? *The Journal of Legal Studies* 29 (2): 913–930.
- Ghauri, P. 2004. Designing and conducting case studies in international business research. *Handbook of Qualitative Research Methods* 1 (1): 109–124.
- Guba, E.G., and Y.S. Lincoln. 1989. *Fourth generation evaluation*. Newbury Park, CA.: Sage.
- Hausman, D.M., and M.S. McPherson. 2008. The philosophical foundations of mainstream normative economics. In *The philosophy of economics: An anthology*, ed. D.M. Hausman, 226–250. New York: Cambridge University Press.
- Hollow, D. 2013. The value for money discourse: risks and opportunities for R4D. Ottawa: IDRC. <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/52733/IDL-52733.pdf?sequence=1&isAllOwed=y>. Accessed 15 Oct 2019.
- IDRC. 2017. The research quality plus (RQ+) assessment instrument. Ottawa: International Development Research Centre. [https://www.idrc.ca/sites/default/files/sp/Documents%20EN/idrc\\_rq\\_assessment\\_instrument\\_september\\_2017.pdf](https://www.idrc.ca/sites/default/files/sp/Documents%20EN/idrc_rq_assessment_instrument_september_2017.pdf). Accessed 30 Sept 2019.
- Independent Commission for Aid Impact. 2011. ICAI's approach to effectiveness and value for money. London: Independent Commission for Aid Impact. <https://icai.independent.gov.uk/wp-content/uploads/ICAI-Approach-to-Effectiveness-and-VFM2.pdf>. Accessed 10 Oct 2019.
- Independent Commission for Aid Impact. 2019. The Newton Fund: A performance review. London: Independent Commission for Aid Impact. <https://icai.independent.gov.uk/wp-content/uploads/The-Newton-Fund.pdf>. Accessed 10 Oct 2019.
- Iwakabe, S., and N. Gazzola. 2009. From single-case studies to practice-based knowledge: Aggregating and synthesizing case studies. *Psychotherapy Research* 19 (4–5): 601–611.
- Keijzer, N., and E. Lundsgaarde. 2018. When 'unintended effects' reveal hidden intentions: Implications of 'mutual benefit' discourses for evaluating development cooperation. *Evaluation and Program Planning* 68: 210–217.
- van de Kerkhof, M., A. Groot, M. Borgstein, and L. Bos-Gorter. 2010. Moving beyond the numbers: A participatory evaluation of sustainability in Dutch agriculture. *Agriculture and Human Values* 27 (3): 307–319.
- King, J. 2019. Combining multiple approaches to valuing in the MUVA female economic empowerment programme. *Evaluation Journal of Australasia* 19 (4): 217–225.
- King, J. 2019b. Evaluation and value for money: development of an approach using explicit evaluative reasoning. PhD thesis, University of Melbourne, Melbourne. <https://minerva-access.unimelb.edu.au/bitstream/handle/11343/225766/PhD%20thesis%20Julian%20King%20July%202019b.pdf?sequence=3&isAllowed=y>. Accessed 30 Sept 2019.
- King, J., K. McKegg, J. Oakden, and N. Wehipeihana. 2013. Rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of Multidisciplinary Evaluation* 9 (21): 11–20.
- King, J. and Oxford Policy Management. 2018. OPM's approach to assessing value for money. Oxford: Oxford Policy Management. <https://www.opml.co.uk/files/Publications/opm-approach-assessing-value-for-money.pdf?noredirect=1>. Accessed 30 Sept 2019.
- Kopp, R.J., A.J. Krupnick, and M. Toman. 1997. *Cost-benefit analysis and regulatory reform: An assessment of the science and the art*. Washington, DC: Resources for the Future.
- Lincoln, Y.S., and E.G. Guba. 1986. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation* 1986 (30): 73–84.
- Marrelli, A. 2007. Collecting data through case studies. *Performance Improvement* 46 (7): 39–44.
- Martens, K. 2018. How program evaluators use and learn to use rubrics to make evaluative reasoning explicit. *Evaluation and Program Planning* 69: 25–32.
- Martin, B.R. 2011. The research excellence framework and the 'impact agenda': Are we creating a Frankenstein monster? *Research Evaluation* 20 (3): 247–254.
- Martin, B.R. and Tang, P. 2007. The benefits of publicly funded research. SPRU Electronic Working Paper Series. Sussex: Science & Technology Policy Research, University of Sussex. <https://www.sussex.ac.uk/webteam/gateway/file.php?name=sewp161.pdf&site=25>. Accessed 30 Sept 2019.
- McLean, R.K.D., and K. Sen. 2019. Making a difference in the real world? A meta-analysis of the quality of use-oriented research using the research quality plus approach. *Research Evaluation* 28 (2): 123–135.
- Næss, P. 2006. Cost-benefit analyses of transportation investments. *Journal of Critical Realism* 5 (1): 32–60.



- National Audit Office. 2016. What is a value for money study? London: National Audit Office. <https://www.nao.org.uk/about-us/wp-content/uploads/sites/12/2016/10/What-is-a-value-for-money-study.pdf>. Accessed 30 Jan 2020.
- Noble, H., and J. Smith. 2015. Issues of validity and reliability in qualitative research. *Evidence-Based Nursing* 18 (2): 34–35.
- Nussbaum, M.C. 2000. The costs of tragedy: Some moral limits of cost-benefit analysis. *The Journal of Legal Studies* 29 (S2): 1005–1036.
- Ofir, Z., T. Schwandt, C. Duggan, and R. McLean. 2016. Research quality plus: A holistic approach to evaluating research. Ottawa: International Development Research Centre. <https://www.idrc.ca/sites/default/files/sp/Documents%20EN/Research-Quality-Plus-A-Holistic-Approach-to-Evaluating-Research.pdf>. Accessed 30 Sept 2019.
- Patton, M.Q. 2002. *Qualitative research & evaluation methods*. Thousand Oaks, CA.: Sage.
- Patton, M.Q. 2015. Sampling, qualitative (purposeful). In *The Blackwell encyclopaedia of sociology*, ed. G. Ritzer, 4006–40007. Oxford: Blackwell.
- Peterson, C., and G. Skolits. 2020. Value for money: A utilization-focused approach to extending the foundation and contribution of economic evaluation. *Evaluation and Program Planning* 80: 101799.
- Ramalingam, B., H. Jones, R. Toussaint, and J. Young. 2008. Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. London: Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf>. Accessed 30 Jan 2020.
- Rethinking Research Collaborative. 2018. Promoting fair and equitable research partnerships to respond to global challenges: recommendations to the UKRI. London: Rethinking Research Collaborative. <https://www.ukri.org/files/international/fair-and-equitable-partnerships-final-report-to-ukri-sept-2018-pdf/>. Accessed 1 May 2020.
- Rogers, P.J. 2008. Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14 (1): 29–48.
- Roling, N. 2011. Pathways for impact: Scientists' different perspectives on agricultural innovation. *International Journal of Agricultural Sustainability* 7 (2): 83–94.
- Rowley, J. 2002. Using case studies in research. *Management Research News* 25 (1): 16–27.
- Scriven, M. 1991. *Evaluation thesaurus*, 4th ed. Newbury Park, CA.: Sage.
- Scriven, M. 2012. The logic of valuing. *New Directions for Evaluation* 2012 (133): 17–28.
- Sen, A. 2000. The discipline of cost-benefit analysis. *The Journal of Legal Studies* 29 (S2): 931–952.
- Shapiro, S.A., and C.H. Schroeder. 2008. Beyond cost-benefit analysis: A pragmatic reorientation. *Harvard Environmental Law Review* 31. <https://papers.ssrn.com/abstract=1087796>. Accessed 31 Jan 2020.
- Shenton, A.K. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information* 22: 63–75.
- Shutt, C. 2015. The politics and practice of value for money. In *The politics of evidence and results in international development: Playing the game to change the rules?*, ed. R. Eyben, I. Guijt, C. Roche, and C. Shutt, 57–78. Rugby: Practical Action.
- Stern, E. 2015. Impact evaluation: A guide for commissioners and managers. London: BOND. [https://www.betterevaluation.org/sites/default/files/60899\\_Impact\\_Evaluation\\_Guide\\_0515.pdf](https://www.betterevaluation.org/sites/default/files/60899_Impact_Evaluation_Guide_0515.pdf). Accessed 14 Jan 2022.
- Stern, N. 2016. Research Excellence Framework (REF) review: Building on success and learning from experience. London: Department for Business, Energy and Industrial Strategy. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/541338/ind-16-9-ref-stern-review.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf). Accessed 15 Oct 2019.
- UK Research and Innovation. 2021. UKRI Official Development Assistance letter 11 March 2021. Swindon: UK Research and Innovation. <https://www.ukri.org/our-work/ukri-oda-letter-11-march-2021/>. Accessed 1 May 2021.
- Vardakoulias, O. 2013. New economics for: Value for Money in international development. London: Nef consulting. [http://bigpushforward.com/wp-content/uploads/2013/02/new-economics-for-VfM-intl-Development\\_FINAL-1.pdf](http://bigpushforward.com/wp-content/uploads/2013/02/new-economics-for-VfM-intl-Development_FINAL-1.pdf). Accessed 2 May 2022.
- Verweij, S., and L.M. Gerrits. 2013. Understanding and researching complexity with qualitative comparative analysis: Evaluating transportation infrastructure projects. *Evaluation* 19 (1): 40–55.
- Walker, E., and B.J. Dewar. 2000. Moving on from interpretivism: An argument for constructivist evaluation. *Journal of Advanced Nursing* 32 (3): 713–720.



---

Yin, R.K. 2018. *Case study research and applications: Design and methods*, 6th ed. Los Angeles, CA.: Sage.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

