Understanding sustainable outcomes in international development

Towards a realist evaluation framework

Feeny, Simon ; Westhorp, Gill; Williams, Emma

RESEARCH ARTICLE

Understanding sustainable outcomes in international development: Towards a realist evaluation framework

Simon Feeny[1] 🄻 | **Gill Westhorp**[2] 🄻 | **Emma Williams**[2] 🄻

[1]RMIT University, Melbourne, Victoria, Australia

[2]Charles Darwin University, Darwin, Northern Territory, Australia

Correspondence
Simon Feeny, School of Economics, Finance and Marketing, RMIT University, 445 Swanston Street, Melbourne 3000, VIC, Australia.
Email: simon.feeny@rmit.edu.au

Abstract

Billions of dollars continue to be provided in foreign aid each year. However, few scholarly studies have examined whether the outcomes from foreign aid interventions are sustained after donor funding has ceased. This paper examines current approaches to assessing this issue before arguing that a realist evaluation approach is ideally suited to understand why and how sustained outcomes are—or are not—achieved. It contributes to the existing literature by presenting three new frameworks to examine the sustainability of outcomes in international development as well as some Context-Mechanism-Outcome statements. Implications for governments, communities, households and donor/implementing organisations are discussed.

KEYWORDS
development outcomes, foreign aid, realist evaluation, sustainability

JEL CLASSIFICATION
F35, F53, I31, O12

1 | INTRODUCTION

The governments of rich industrialised countries have provided several trillion U.S. dollars in foreign aid to low- and middle-income countries (LMICs) since the Second World War (Tierney et al., 2011). In 2019 alone, Organisation for

Economic Co-operation and Development (OECD) member countries provided an estimated USD 152.8 billion in Official Development Assistance (ODA), the most commonly used measure of foreign aid, to promote development globally (OECD, 2020). Whether or not aid 'works' has been hotly contested over the decades (see, for example, Sachs, 2005; Easterly, 2007; Riddell, 2008; Moyo, 2009). While much work has been carried out to assess the outcomes of foreign aid, few scholarly studies have examined the extent to which outcomes, positive or negative, are sustained. Even less is known about how and why sustained outcomes are—or are not—achieved. Further, while there is widespread agreement that the positive outcomes of development interventions need to be sustained, there is no established framework to inform development actors (communities and governments in LMICs, non-government organisations, funders and the like) about how this may best be achieved or to underpin evaluations of sustainability. This represents a serious constraint in the quest to improve well-being in developing countries. Without sustained positive benefits, development interventions will do little to support the achievement of the United Nations' Sustainable Development Goals (SDGs) by 2030.

Even how sustainability should be defined or conceptualised is yet to be agreed (Moore et al., 2017; Taylor, 2014). The sustainability of development intervention outcomes should not be confused with the concept of sustainable development. The latter is often defined as development that meets the needs of the present without compromising the ability of future generations to meet their own needs (Brundtland Report, 1987). The former refers to the outcomes of specific programmes and whether or not they are sustained. A second distinction is between sustainability of the outcomes achieved by interventions and sustainability of the interventions themselves. In some circumstances, programming will need to be sustained in order for outcomes to be sustained; other types of outcomes, and other pathways to outcomes, may not require sustained programming.

This paper has been written as part of a research programme investigating the sustainability of outcomes in international development, and we will, in general, use this term ('sustainability'). Our programme begins from the assumption that theories or frameworks for sustainability should be grounded in research about what has, in fact, been sustained. Our definition of sustainability therefore builds on our definition of 'sustained outcomes'. We define sustained outcomes as the persistence of net benefits, directly or indirectly attributable to an international development intervention, after external funding has ceased. We exclude from this definition aspects of the OECD DAC definition of 'sustainability' as being outcomes that are 'likely to continue' because it is open to considerably greater subjectivity and debate (see OECD, 2021). We also exclude the idea, used in at least one NGO's definition, of sustainability being 'the ability to maintain' (WVI, 2016). While capacity is certainly one key determinant of whether or not outcomes will be sustained, that capacity may or may not be directed to doing so. The circumstances in which it is, or is not, should form part of the investigation. Rather, we seek to determine what *is* sustained, or not, under what conditions and how and why (or why not). Note that the term 'net benefits' captures both intended and unintended outcomes from an intervention; includes new outcomes that may emerge after the project but as a result of project work; and implies that both benefits and disbenefits are taken into account and need not imply that all costs and benefits can be converted to dollar terms. This definition captures the elements common to those used by some international development agencies (see, for example, USAID, 2020; DFAT, 2021).

To make the distinction between the sustainability of outcomes versus the sustainability of interventions, it is useful to consider the following example. Health interventions in early life may contribute to improved outcomes in later life for the children who received them (or whose parents did so). These outcomes can include improved health, educational attainment, mental health, economic productivity (although only for male participants in one Guatemalan study) and reduced violent behaviour (e.g., Gertler et al., 2014; Hoddinott et al., 2008; Maluccio et al., 2009; Stein et al., 2008; Walker et al., 2011). To the extent that the health interventions worked by changing parent behaviours, parents may (in the right circumstances) continue those behaviours with subsequent children, so here too, outcomes may be sustained. However, to the extent that the intervention relied on the provision of health services tailored to the particular needs of individual children, benefits to subsequent generations of infants rely on those health services

continuing to be provided. If the intervention ceases when donor funding ceases, benefits to those subsequent generations are unlikely to be achieved and cannot therefore be sustained. Different outcome patterns will therefore be evident for the sustainability of different kinds of outcomes, depending on the processes and mechanisms that generate the outcomes.

In other cases, sustaining benefits may not require the original intervention to be sustained but may require different interventions to be introduced. If the original intervention increased school enrolments, subsequent interventions (for example, increasing the number of qualified teachers) may be required to sustain the same level of service for the higher number of enrolments.

Regardless of whether interventions need to be maintained or new interventions need to be introduced, benefits (and disbenefits) are likely to be sustained in some contexts but not others. We therefore argue that understanding sustainability lends itself to a realist evaluation approach. Realist evaluation is a form of theory-driven evaluation that seeks to identify what works, for whom, in what respects, to what extent, in what contexts and how, typically using Context-Mechanism-Outcome (CMO) statements to postulate and refine hypotheses (Pawson & Tilley, 1997; Westhorp, 2014). Further details are provided below.

We contribute to the existing literature by presenting three new frameworks to examine the sustainability of outcomes in international development. The frameworks were developed as a product of previous evaluations and incorporated in final evaluation reports provided to World Vision but have not otherwise been published. The authors are unaware of any use made of the frameworks thus far, but they provide starting points for the development of realist theory. In this paper, we use them to postulate some initial CMO statements about outcome sustainability. The examination of sustained outcomes from a realist perspective is still in its infancy, and the theories presented here are intended to be tested and refined through further evaluations.

The remainder of this paper is structured as follows. Section 2 summarises existing approaches to assessing sustainability in the international development sector while Section 3 makes the case for the adoption of a realist approach. Section 4 provides existing theoretical frameworks for assessing sustainability. Section 5 provides some new frameworks that were developed from realist evaluations and could underpin future evaluations of sustainability of outcomes. Section 6 discusses the implications of these frameworks for the roles and contributions of four different sectors to sustained benefits: individuals or households; communities; governments; and the organisations that implement the interventions. Finally, Section 7 concludes by noting how the frameworks can be used at different stages of intervention development and implementation.

2 | APPROACHES TO ASSESSING SUSTAINABILITY OF OUTCOMES

There is a sparse scholarly literature examining the sustainability of development interventions. Recent examples include Fowler (2013), Cekan (2016) and Ishola and Cekan (2019). Studies often lack an underlying conceptual framework to guide their approach and analysis. The literature presents two approaches. The first, undertaken by multilateral institutions including the World Bank and the Asian Development Bank (ADB), is applied at the end of an intervention's implementation period but before further time has elapsed. Independent assessors, at the time of evaluation, rate how likely it is that the benefits from an aid project will be sustained beyond the life of the project, given their consideration of different risk factors.

More specifically, in evaluations of World Bank projects since 2007, the Independent Evaluation Group (IEG) of the World Bank assesses what is termed the 'risk to development outcome'. This is defined as 'the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized)' (IEG, 2015, p. 16). The risk is assessed using a Likert scale (negligible to low, low, moderate, significant or high). The authors' analysis of the 2020 release of World Bank project performance data (World Bank, 2020) reveals that 47% of 3192 evaluated projects faced a 'significant' or 'high' risk of outcomes (or expected outcomes) not being maintained (or realised).

Sustainability is also an important component of the project performance assessments conducted by the Independent Evaluation Department (IED) of the ADB.[1] Here, sustainability relates to 'the likelihood of the changes brought about by the project being continued in the long term' (IED, 2020, p. xi). The IED rates sustainability as most likely, likely, less than likely and unlikely. In its most recent Annual Evaluation Review, the IED found that 38% of ADB sovereign operations rated sustainability as 'less than likely' or 'unlikely' in 2017–2019. The report concluded that sustainability is still the weakest element in the performance of ADB sovereign operations and that there is a need to ramp up efforts to address the issues causing weak sustainability of project outcomes (IED, 2020).

The second approach more directly assesses sustainment of outcomes. It uses ex post evaluations, which are conducted after a project has been completed. There is no accepted time period following project completion at which an ex post evaluation should be conducted, but it is typically 2–10 years. Both quantitative and qualitative data are usually collected for these evaluations. These evaluations usually focus on outcomes and often identify factors contributing to them. However, they fall short of realist work because they usually do not use programme theory, identify mechanisms or examine patterns of outcomes for subgroups or different contexts. There have been large numbers of ex post evaluations. Here, we review three recent ex post evaluations conducted by large international development organisations.[2]

A good example of the ex post approach is provided by an Inter-American Development Bank assessment of the sustainability of 100 rural water systems that it financed in Paraguay. Specifically, it examined the operational status of the water systems 8–10 years after they were installed as well as the quality of the service provided. The study found that 40% of the systems were not working at any moment in time and that 10 years after installation, 25% of the systems did not work at all.[3]

USAID (2020) conducted ex post evaluations of four rural and two urban Water, Sanitation and Hygiene (WASH) activities in six countries: Ethiopia, India, Indonesia, Madagascar, Mozambique and Senegal. The evaluations were conducted 3–10 years after the USAID-funded activities ended. The report found that, ex post, the functionality of water points ranged from 44% to 65% in predominantly rural settings in Ethiopia, Mozambique and Senegal. Access to water had increased in Indonesia and had remained the same in one site in India while increasing in another. Access to soap ranged from 0% in Ethiopia to 31% in Senegal. Latrine use had fallen in the case of Madagascar, and the practice of open defecation continued in Ethiopia, Madagascar and Senegal.

In 2015, Catholic Relief Services conducted an evaluation of a multi-partnered food security project implanted in Niger during the period 2006 to 2012 (CRS, 2016). Findings were viewed as being very positive with 80% of activities reported to have become self-sustained. Food security had increased, women's income improved and over 90% of households reported improvements in the health and well-being. Moreover, new community innovations emerged that community members attributed to the project such as collective funds to pay for cleaning of a new health centre, community-imposed sanctions for births occurring outside of health centres and the monitoring of savings from well water sales (CRS, 2016).

Some studies assess sustainability using randomised control trials (RCTs) for ex post evaluations. An RCT randomly assigns a development intervention to some members of a population known as a 'treatment' group, and outcomes are compared to those for members that did not receive the intervention, known as a control group. Given the random allocation of the intervention, any difference in outcomes between the treatment and control groups is assumed to be caused by the intervention. Follow-ups are sometimes possible with these RCTs.

This approach has been used to examine the sustainability of outcomes from microfinance programmes. For example, de Mel et al. (2012) found that enterprise survival rates and profits were higher for male-owned Sri Lankan

---

[1]The other components of assessed performance are relevance (relating to the extent that project objectives aligned with the priorities of major stakeholders and whether project designs were appropriate), effectiveness (relating to the extent to which planned outputs and outcomes were actually achieved) and efficiency (relating to whether the resources used to achieve the outcomes was optimal) (see IED, 2020).

[2]Recent published studies that focus on ex post evaluation include Bishop (2018), Cisilino et al. (2019), Faure et al. (2020) and Gómez-Lobo (2020).

[3]Clarke, Feeny and Donnelly (2014), examining WASH infrastructure in the Pacific region 3–5 years after projects had completed, found only one of 21 projects in which benefits had been sustained at the same level or better than when the project completed. Almost half of the water systems established were no longer operating as intended.

microenterprises, 5 years after they were given a cash grant, compared to those who did not receive the grant. There were no impacts for female-owned businesses—begging the question why intervention outcomes were sustained for men but not for women. Blattman et al. (2014, 2018) found that an RCT that provided some women cash grants and basic business training in northern Uganda led to large income gains 4 and 9 years after their receipt. There are similar findings for cash grants and in-kind capital to entrepreneurs in Ghana and India for up to 5 years after the intervention (Fafchamps et al., 2014; Rigol et al., 2017). Despite their popularity, RCTs are very expensive to run, and they are usually unable to identify the reasons why the outcomes of an intervention are, or are not, sustainable.

3 | A REALIST APPROACH TO ASSESSING SUSTAINABLE OUTCOMES IN INTERNATIONAL DEVELOPMENT

This paper argues that a realist approach is ideally suited to evaluating sustainable outcomes in international development. Realist evaluation is a theory-based approach that seeks to explain how and why things work in different contexts rather than focusing on average outcomes. The standard question in a realist evaluation is 'what works, for whom, in what respects, to what extent, in what circumstances and how?' (Pawson & Tilley, 1997; Westhorp, 2014).

We identify four main reasons why realist evaluation is fit-for-purpose. Firstly, it is based on programme theories that are (continually) developed through an iterative process. Initial proposed theories are tested, refined or refuted during the evaluation. The hypotheses that are developed from programme theories are called CMO statements. They explain how (causal) Mechanisms (M) interact with characteristics of a Context (C) to generate Outcomes (O). It is important to note that mechanisms are not the programme interventions themselves but (usually) invisible changes in people's reasoning, capacity and motivation that might arise from, or in response to, the resources or opportunities provided by the intervention (see Dalkin et al., 2015; Westhorp, 2018). Programme theories and CMOs can be specified at different levels of abstraction. At the most granular level, they will apply to very specific elements of a programme, while middle range theories can apply across different kinds of programmes (Pawson & Tilley, 1997; Punton et al., 2020) or to programmes in different settings. Given the paucity of established programme theories that seek to establish how sustained outcomes in international development are achieved when they are, and why they are not in other cases, a realist approach will (amongst its other uses) contribute to filling this important gap.

Secondly, community-led development interventions occur in complex social systems. Programme interventions cannot be expected to work in the same way in these systems. A realist approach is able to identify different ways in which interventions work as well as the conditions required for them to work.[4] Development actors will therefore benefit from rich, nuanced findings and an understanding of whether, where and how their development interventions are likely to be portable to other contexts.[5] Thirdly, sustainable outcomes in international development will be largely determined by people, and people are front and centre in the realist approach. The approach recognises that it is people who create change, and for interventions to work, participants must be motivated, engaged and have the capacity to act in an enabling environment (Mukova-Moses, 2017; Pawson & Tilley, 1997).

Finally, realist evaluation is method-neutral in its approach, and both quantitative and qualitative data are typically collected and analysed to test the programme theories. This means that whatever data type was used for an end-of-programme evaluation can be incorporated in an evaluation of whether outcomes were sustained; that multiple data types can be brought to bear in identifying and testing theories about how and why sustainability was or was not generated; and that data can be interrogated in various ways to support triangulation and confidence in findings.

This adoption of a realist evaluation approach represents an important departure from the evaluation techniques currently used in international development. RCTs, discussed in the introduction, have sometimes been viewed as

---

[4]There is a growing literature using realist evaluation in the field of international development. Recent examples in the field of health include Marchal et al. (2010); Graham et al. (2018); and Dossou et al. (2021). These studies identify the different ways in which programmes work for different people.
[5]This addresses the critique of external validity that applies to RCTs and other evaluation approaches (whereby the results from an evaluation cannot necessarily be generalised to other contexts).

the gold standard in evaluation. However, the limitations of RCTs are well established, particularly for interventions in complex settings that work differently for different people (see, for example, Deaton & Cartwright, 2018; Ravallion, 2020; Tilley & Westhorp, 2020). Nobel Prize winning economist Angus Deaton notes that 'RCTs are informative about the mean of the treatment effects but do not identify other features of the distribution' (Deaton, 2010, p. 439). Deaton also argues that 'the analysis of projects needs to be refocused toward the investigation of potentially generalisable mechanisms that explain why and in what contexts projects can be expected to work' (Deaton, 2010, p. 426).[6]

We seek to build upon some recent, realist contributions in the literature. Douthwaite et al. (2017) use realist synthesis and other techniques to develop design principles for monitoring and evaluation for complex interventions. Owusu-Addo et al. (2020) conduct realist evaluation to develop middle-range theory that explains how cash transfers influence the social determinants of health, while Hoffecker (2021) draws on realist evaluation methods to develop a middle range model of inclusive innovation processes within complex systems to inform programme theory-building, implementation and evaluation. Our programme takes this further by applying realist approaches to the sustainability of outcomes in international development.

4 | THEORETICAL FRAMEWORKS FOR SUSTAINABILITY

Assessments of sustainability will continue to provide limited insights for development actors in the absence of programme theories for sustainability—that is, to underpin the design, monitoring and evaluation of sustainability of the outcomes from their development interventions. The issue of sustainability must be considered from project inception forward and the factors that are likely to be important for benefits to continue monitored closely throughout the project, including transition to local ownership.

Many organisations have their own frameworks for assessing sustainability; here, we present two that can inform the achievement of sustained outcomes from large multisector development programmes.[7] The first is a framework developed by The Springfield Centre (2014) for a *Making Markets Work for the Poor* programme, funded by the Swiss Agency for Development and Cooperation (SDC) and the U.K.'s Department for International Development (DfID). The framework seeks to identify what functions and rules need to be in place for the programme to continue benefitting the poor after the intervention has ended as well as who performs those functions. This includes the identification of who is currently doing what and paying for it and who will do what and who will pay for it in the future (see Figure A1 in the appendix). The Springfield Centre (2014) report also poses key questions that need to be answered in order to ascertain whether the programme is sustainable when donor funding ceases. These questions include, if you were to leave now, would partners adopt and adapt? Would they build on changes or return to their old ways of working? Would the system be able to support the changes that occurred, or would the benefits depend on too few people, firms or organisations? (see Figure A2 in the appendix). Clearly, these questions are predictive in nature.

A second framework is provided by Rogers and Coates (2015) in undertaking a four-country study of sustainability and exit strategies for food assistance projects. They identified factors that were expected to lead to sustainability, defined as continued benefit after the end of a project. The framework is based on the programme sustaining three intermediate outcomes: (i) the creation or strengthening of service delivery mechanisms, (ii) beneficiary access to services and (iii) beneficiary demand for services. The study team hypothesised that sustaining these outcomes

---

[6]Note that experimentalists and realists differ in their approach to causality. Experimentalists identify 'successionist' causality where a cause operates constantly and regularly to produce an effect. In these cases, if A 'causes' B, what the cause actually does to produce the effect is not revealed. Through the identification of 'mechanisms', realists identify 'generative' causality and ascertain how a causal association actually comes about (DFID, 2012; Westhorp, 2014, 2018).

[7]Other related resources include a manual to planning and measuring sustainability in health programs (Sarriot et al., 2008) and a sustainability framework and assessment tool to help build the capacity for maintaining public health programmes (Schell et al., 2013). GHPC (2016) and WVI (2016) identify the factors that are likely to be important for outcomes to be sustained.

requires (i) a sustained source of resources; (ii) sustained technical and managerial capacity; (3) sustained motivation and incentives that do not rely on programme inputs; and often (4) sustained linkages to other organisations that can promote sustainability by augmenting resources, refreshing capacity and motivating service providers and beneficiaries (see Figure A3).

Theoretical frameworks can assist donors and implementation agencies to devise, monitor and evaluate particular aspects of the sustainability of outcomes and/or development interventions. How to identify sustained outcomes, how they were achieved, for whom and in what circumstances are all important questions. The two frameworks above answer elements of these questions, but the predictive questions in the Springfield Centre report—while useful for planning—cannot assess whether, how or why outcomes were sustained. Similarly, the framework provided by Rogers and Coates does not address how or why the elements in it are themselves likely to be sustained or the contexts in which they are more, or less, likely to be so. A different approach is needed to answer these questions.

5 | TESTABLE FRAMEWORKS FOR SUSTAINABILITY

In this section, we demonstrate how a framework can be used to generate testable hypotheses for realist investigation of how and why outcomes are sustained—or not—in different contexts, using CMO statements. Each of the frameworks below, while not realist in their structure, was developed by the current authors as a product of research or evaluation projects undertaken for World Vision International. Two of the three frameworks were used to summarise findings or insights arising from the evaluations. The third was used to inform the development of evaluation tools and refined on the basis of the findings; it has been simplified for use in this new work. The frameworks are not intended to be specific to World Vision as an organisation; they are developed 'at a middle level of abstraction' that should apply across types of programmes and organisational providers. However, given World Vision's approaches, it is possible that they may be most easily transferred to community-based (or 'community led') approaches to development.

5.1 | Pathways to sustainability

The first framework (Framework 1 below) was developed as part of a research project investigating the effectiveness of World Vision's community-based approach to child-focused development, delivered through Area Development Programs (ADPs) (Feeny et al., 2017). One of the questions for that evaluation was 'How do WV programmes enable communities to sustain improvements in child wellbeing?' Here, the overarching outcome of interest is 'improvements in child wellbeing'. Considered from the perspective of 'sustained benefits', this could include sustaining benefits to those involved at the time of the intervention and/or similar benefits accruing to subsequent generations of children. Benefits could be of many types, including improved health or improved education outcomes.

The framework provides a typology of pathways to sustainability of outcomes. It posits a total of nine pathways. In column 1, there are three pathways at national or regional levels, likely to operate primarily through governments: regulatory change, increased service system capacity and mainstreaming of services. In these pathways, benefits may scale or spread over time and/or be maintained at a regional or national level because authority holders have accepted responsibility for sustained outcomes.

In the second column are three pathways posited to operate at community level: maintenance by social infrastructure, maintenance of physical infrastructure and new social norms. In these pathways, benefits may scale to additional members of the communities involved in the programme but are less likely (or would be slower) to transfer to other communities. In the final column are three pathways that operate at household or individual level: maintenance of behaviour, diffusion of behaviour, and maintenance of individual benefit. In this column, changes may be maintained amongst participating families and their immediate networks but may not scale or spread further. Widespread diffusion is unlikely because diffusion generally occurs only within close social networks such as kinship

groups, friendship groups and membership of the same organisations or groups; physical proximity (being neighbours) is not usually enough (Kaaria et al., 2004; Kiptot et al., 2006).

Framework 1: Pathways to sustainability

1 Regional/national level	2 Community level	3 Individual/family level
Government regulation: Outcomes are sustained through changes to legislation or regulations at national, district or local levels. Responsibility has been accepted by appropriate authority holders for continuation, maintenance and monitoring of regulations or programmes that sustain outcomes.	Social infrastructure: Outcomes are sustained because services or activities are maintained by local organisations or committees, with or without support from external organisations or governments.	Behaviour: Outcomes are sustained because behaviours learned and adopted by community members have become ingrained. Improvements in well-being are largely limited to those who participated in the programmes and changed their behaviours at the time.
Capacity of service systems: Outcomes are sustained through improvements to funding levels, policies or procedures, capabilities and/or attitudes of actors in service delivery systems. Responsibility for maintenance of the change is accepted by the authority holders in the service delivery system. The original programme may not be maintained.	Material infrastructure: Outcomes are sustained because physical infrastructure that was built during initial programmes is maintained after the initial intervention ends, such that the benefits of the infrastructure continue to accrue to community members.	Diffusion of behaviours: Outcomes are sustained because participants deliberately pass forward knowledge or skills learned through their own participation to others or because non-participants observe benefits and decide to act in similar ways in the interests of securing those benefits.
Mainstreaming: Outcomes are sustained because projects or programmes initiated by one organisation are mainstreamed. Train the trainer approaches provide the possibility of increasing scale, capacity and/or quality improvement over time.	Cultural or normative change: Outcomes are sustained because norms or cultural practices that undermine them have been changed or because relationships between groups have been improved such that risks of negative situations or behaviours are decreased.	Maintenance of benefits: Outcomes are sustained simply in the sense that benefits achieved during the programme continue to exist, whether or not behaviour changes are maintained.

Source: Adapted from Feeny et al. (2017), p 25.

It is, of course, possible that there will be interactions between these different kinds of sustainability and that some may operate as mutually reinforcing feedback loops. For example, changes in cultural norms may support regulation or policy change, and regulation or policy change may support cultural or normative change. It is also possible that benefits at one level of the system (government, community or family) may over time affect other levels of the system. However, it does not automatically follow that outcomes for individuals will translate to outcomes for communities, any more than it follows that outcomes for communities or changes by governments or service systems will translate to benefits for all families.

A significant shortfall of this framework is that it excludes—or at best, lacks transparency about—interventions which are expected to operate through the private sector—for example, value chain development. These may require development of private sector infrastructure, capacity building for different roles along the value chain and sufficient return on investment for the system to be self-sustaining. Resilience to climate shocks and market shocks will also be required for the system to self-sustain. It is likely that congruence between the implementation model and the intended final system is required. For example, in order to test whether an intervention or an innovation can survive

TABLE 1 Initial CMOs for pathways to sustainability

Context	Mechanism	Outcome
Programmes conducted by NGO or other organisations demonstrate positive effects in relation to a policy priority for the government. The level of government (national, state or local) has the mandate for achievement of outcomes in relation to the issue. The level of government has adequate financial and human resources to implement/administer the policy or regulation. There is sufficient community/population level support for the outcome of interest.	Authority holders are convinced by evidence that the legislation/regulation will achieve intended outcomes; believe legislation/regulation is the appropriate strategy and is economically and politically viable; and accept responsibility for implementation, maintenance and monitoring of legislation or regulations.	Legislation or regulations are introduced at national, district or local levels.
Legislation or regulations are introduced	Compliance with legislation or regulations	Behaviour changes that sustain outcomes

in the market, both its effectiveness and the capacity and willingness of the market to pay for it must be tested and demonstrated. A second shortfall of the framework, at least for realist purposes, is that it does not describe the causal mechanisms that underpin the pathways, whether (and if so how) they differ across pathways, or what is required in the context for those mechanisms to operate.

In Table 1 below, we demonstrate how 'initial rough' (Pawson, 2002) CMOs could be developed from the framework, for testing in an evaluation. The example relates to the first pathway in the first column of Framework 1—the introduction of legislation or regulations. Two CMOs are identified: the first relating to the introduction of the legislation and the second to its contribution to sustained outcomes. In the first row, the 'context' column identifies a series of conditions that are expected to apply by the time the initial programme (that is, the programme whose outcomes should be sustained) concludes. The 'mechanism' column refers to 'reasoning' on the part of authority holders in response to those conditions, and the outcome is introduction of legislation or regulations. The second CMO starts where the first leaves off. The introduction of the legislation or regulation now becomes the context in which a new mechanism (compliance with the requirements of the legislation or regulation) fires, resulting in changed practices or behaviours that sustain the outcome(s) of interest.

CMOs could be developed for each of the nine pathways. Even without them, though, Framework 1 may provide a starting point for realist questions to be framed. For example, in what contexts are governments most (and least) likely to change regulatory frameworks, expand service system capacity, or mainstream services, and why?

A first step towards such use of the framework was demonstrated in a later realist-informed research project for World Vision (Westhorp et al., 2017). That project examined (amongst other things) whether and how (a) fidelity to programme models and (b) adaptation of models to context contributed to sustainability. The nine pathways were used as a reporting framework, examining the contexts in which particular pathways appeared or did not. So, for example, four contexts were identified which appeared to make the government regulation pathway less likely: poorer countries, which lacked the resources to implement national programmes or enforce national legislation; sustained conflict and instability (e.g., in the Democratic Republic of Congo); emergencies (e.g., the Ebola crisis); and lower levels of community awareness and support (e.g., in Cambodia) (Westhorp et al., 2017, p. 81).[8] Regulatory change appeared more likely in relatively stable countries, with somewhat better resources and capacity and where coalitions of organisations operated together to advocate for change.

Programmes (or components of them) were more likely to be mainstreamed where the country had adequate and relatively stable service delivery systems; the programme model was consistent with existing service delivery

---

[8]Note that all these contexts were relevant at the time of the report—but contexts in countries change over time.

TABLE 2 Example context, mechanism and outcomes for sustainability

Context	Mechanism	Intermediate outcome (pathway to sustainability)
Poorer countries, insufficient resources for implementation or enforcement	Issue not regarded as highest priority. Governments unwilling to create risks to governance (e.g., lack of trust) by establishing frameworks that cannot be enforced.[a]	No changes to legislative, policy or regulatory frameworks
Conflict and instability Power struggles	Lack of priority afforded to the target issue. Lack of agreement about directions or strategies.	

[a]See OECD (2018).

models and addressed a priority issue for the government, and the functions could be subsumed into existing staff (or volunteer) roles (ibid p. 83).

These findings have informed the contextual factors identified in the first example CMO above. This, in turn, demonstrates how findings from series of studies can be used to iteratively refine realist hypotheses.

These findings relate only to context; they do not tell us what the mechanisms which operate in those contexts are. It is entirely possible (and perhaps likely) that the mechanisms in the poorest countries are somewhat different from those in middle income countries, for example, and that the mechanisms affected by high levels of conflict and instability are different from those affected by resource availability. However, even this much information can provide a basis for hypothesising CMOs for sustainability, which could be investigated in future research, as Table 2 below demonstrates.

5.2 | Resources for sustainability

The elephant in the room in relation to sustainability is resourcing. International development agencies act in countries and sites that lack the resources to meet their own needs, and many aim to build their capacity to do so. However, their interventions are necessarily limited in geographic spread, timescale and cost. Further, many international development agencies focus on the poorest and most vulnerable—by definition, those with the least capacity to sustain activities and outcomes and for whom it might be expected that more intensive, longer lasting and more expensive interventions may be required.

Communities, however poor, do have their own resources on which to draw. Development agencies neither can, nor should, go on providing all the kinds of resources that are required to achieve sustained outcomes. We argue, however, for clarity about the *kinds* of resources that are required for sustainability, the ways that those resources are used, and the risks that are attached to them.

The second framework addresses precisely these questions. It was developed as part of the second research project described above, investigating how programme fidelity and adaptation to context could contribute to programme effectiveness and sustainability (Westhorp et al., 2017). One of the research questions for this project was 'In what circumstances and to what extent have our models[9] been owned by the communities and local partners? In what circumstances have they been able to (or are likely to) continue functioning without World Vision's support?'. Here, the outcome was conceptualised as continued functioning of a programme, with the intent that it should continue to provide benefits to programme beneficiaries. The primary mechanism expected to contribute to this continuation was named in the first part of the question: community and local partner 'ownership' of the programme model (potentially a vexed issue, given that communities are not homogenous and different subgroups within communities may or may not feel a sense of ownership). Regardless, 'ownership' will only result in programme continuation if the

---

[9]World Vision uses the term 'model' to refer to specific ways of implementing projects or services. Models are expected to be evidence-based, and fidelity to the model is thus expected to provide consistency with evidence of 'what works'. Elsewhere in this paper, we have used the more common term 'programmes' and adopt the term 'frameworks' to refer to the analytic models presented here.

resources are available to continue it and so one of the analyses undertaken related to the types of resources required. The four different types of resources required for sustained outcomes are provided in Framework 2 below.

Framework 2: Resources for sustainability

Financial resources	Programme resources
• _Pay salaries/honoraria_ • _Pay transport & other costs_ • _Buy consumable resources_ Provided by development organisation during the programme Require continual replenishment/input both during and after development organisation involvement.	• _Models_ • _Curricula and training materials_ • _Other materials_ Initially external to the community, imported and contextualised Can be retained for re-use in the community but requires adequate expertise to do so.
Resources created by the programme	Local resources mobilised by the programme
• _Social infrastructure_ • _Material infrastructure_ Use both imported (know-how, facilitation, costs) and local (human, relationships) resources to develop Retained in community but require active maintenance (financial & human resources) Can generate/mobilise other resources	• _Existing social infrastructure_ • _Existing powers, authority, capacities_ • _Existing norms, motivations_ Pre-existed the programme but 'put to work' in the service of the programme objectives Retained in the community Can be 'distracted' to other purposes

In summary, the framework suggests that development agencies use financial, programme and existing community resources to create social and material infrastructure. The intent is that the social infrastructure should then use its enhanced capacity, along with retained programme and local resources, to sustain activity and benefits.[10] However, this is only possible where communities also have, or have access to, sufficient social and financial resources to sustain that activity. That includes sufficient funds to pay operating expenses at the very least, and in many programmes, funding to pay salaries or honoraria as well. It seems unlikely that activities and therefore outcomes will be sustained unless financial resource requirements are also sustained.

Intervention agencies play a critical role here, both in providing—at least for some period of time—new resources and in mobilising local staff and volunteers. The capacity of human resources to achieve sustained outcomes can be increased by training and potentially by strengthening aspects of social infrastructure. However, human resources in poor communities are a finite resource, with multiple calls upon them. Turnover in personnel may undermine sustainability because trained personnel move on to other possibilities and either are not replaced or are replaced by people who did not take part in capacity building.

Again, we provide an example of how the framework might be developed into CMOs for testing (Table 3). For simplicity, we choose the example of material infrastructure built by a programme (think building of dams, schools or health clinics). In this case, we have provided one 'positive' CMO, in which outcomes are more likely to be sustained (the second CMO), and one 'negative' CMO in which outcomes are likely to wane.

5.3 | Fidelity and contextualisation

The third framework draws from the same research project as the second (Westhorp et al., 2017). The framework was initially developed from a half-day workshop with World Vision UK and World Vision International staff and reflected their assumptions about how 'fidelity to evidence-based programming' and 'being adapted to context' would contribute to sustainability. These two features are policy requirements for World Vision programmes.

---

[10]'Cognitive' infrastructure is another type that could be included in the frameworks. Uphoff and Wijayaratna (2000) demonstrate the importance of social capital in for sustained outcomes in farming in Sri Lanka.

TABLE 3 Initial CMOs for resource generation and use

Context	Mechanism	Outcome
Infrastructure is built in response to need identified by community leaders (not whole community). Funding and labour for building is provided externally (e.g., by NGO) Unclear who is responsible for maintenance.	Low sense of ownership of the infrastructure. Community believes the problem has been addressed.	Low priority afforded to maintenance of infrastructure. Community resources allocated to other priorities. The contribution of the infrastructure to well-being wanes over time.
Infrastructure is built in response to need identified in community wide consultation. Clear agreements are established in advance about responsibilities and plans for maintenance. Capacity building provided re maintenance if required. Community is engaged in providing labour and materials and applying for funding if required.	High sense of community ownership helps to hold those responsible for maintenance accountable.	Infrastructure is maintained, and its contribution to outcomes is sustained.

Note that these features are not specific to particular programmes but can apply to programmes in many domains; they are therefore further examples of 'middle range theory'. Middle range theory, originally described by Merton (1968), is specific enough that hypotheses can be drawn from it to test and general enough to apply across programmes or contexts (Pawson & Tilley, 1997). It is a key feature of realist research and evaluation and contributes to the portability of findings.

Some data were collected in relation to the framework during the evaluation, although not all aspects of it were tested. The framework was then somewhat refined using the data collected. The refined framework, as incorporated in the evaluation report, is represented and described further below. This process of partial testing and iterative refinement is consistent with realist work (and indeed with research in general).

Framework 3: Fidelity, contextualisation and sustainability

Revised theory diagram: FIDELITY, CONTEXTUALISATION AND SUSTAINABILITY
Focus question: How do fidelity and contextualisation contribute to sustainability?

WILEY
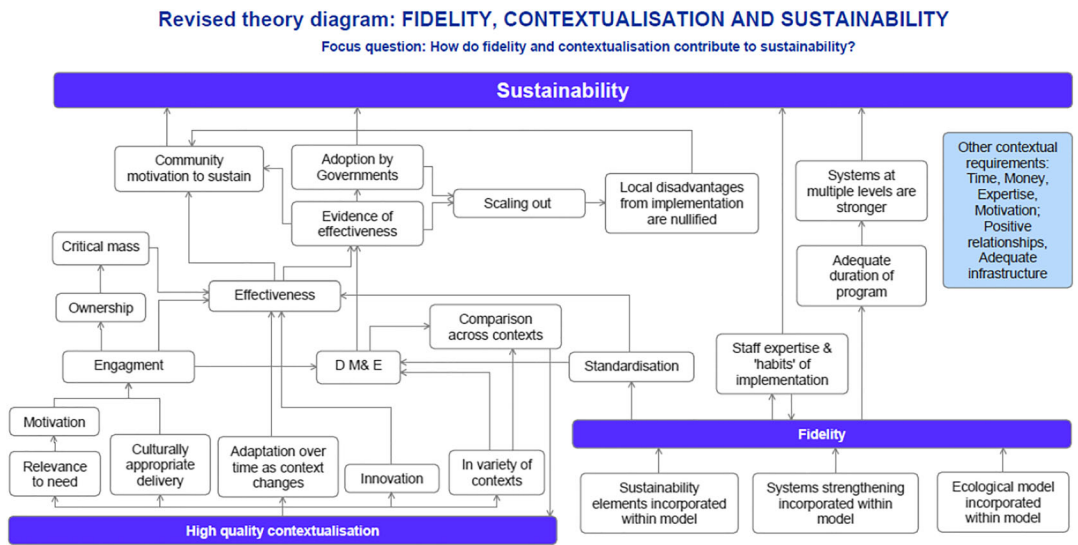
The two main features being examined (high-quality contextualisation and fidelity to evidence-based programming) appear in blue boxes at the base of the framework.

High-quality contextualisation was expected to contribute to sustainability through a number of channels. Firstly, contextualisation should ensure relevance to need, which contributes to 'motivation' and thence to engagement, and through this to a sense of ownership of the programme. Secondly, culturally appropriate delivery also contributes to engagement. Thirdly, innovation as a result of adapting to the context could contribute to effectiveness in that context but could also potentially contribute to replication or adaptation of those innovations in other settings. Fourthly, adaptation over time in response to changes in the context changes should contribute to ongoing effectiveness. Effectiveness, in turn, could contribute to communities' and governments' motivation to maintain the model. Monitoring and evaluation both check whether adaptations are effective and, where they are, provide evidence to inform maintenance and/or scaling out of the programme. Finally, it was expected that monitoring and evaluation across contexts would enable comparisons across those contexts, which should inform decision-making about programme design and implementation; whether to continue or extend use of adaptations; the need for further adaptation; in which sorts of contexts the framework should and should not be implemented; and so on. The evaluation found evidence for some of these elements of the programme theory and identified areas for further development.

Capacity building was incorporated as one 'sustainability element' in the programmes being evaluated (other elements were identified but are not addressed in this framework). Fidelity to capacity building, system strengthening and working across multiple levels of systems were expected to contribute to stronger systems, which, in turn, should contribute to sustainability. Fidelity to the framework across sites could also support effective learning about what works in what contexts for whom. That learning could then be used to refine the framework, potentially introducing greater variability in programme design (through contextualisation) but also contributing to effectiveness. Fidelity could also contribute to staff expertise and to effective 'habits' of implementation, which operate in a feedback loop to maintain ongoing fidelity to the framework.

The framework combines many pathways and components. This is consistent with the complexity of development programming and the complexity of achieving sustained outcomes. It is also consistent with the complexity-theory basis of realist evaluation (Westhorp, 2012, 2013). Consequently, many CMOs could be developed from it. Here, we demonstrate just two in Table 4 below—one positive and one negative—related to the notion that fidelity to the programme model could contribute to learning about the programme and thus to programme adaptation and effectiveness. This example demonstrates the value of realist approaches in addressing organisational contexts—aspects of the implementing organisation that affect whether, how and for whom programmes are effective.

6 | IMPLICATIONS OF THE FRAMEWORKS

In the next section of this article, we consider the implications of these frameworks for different levels of systems: individuals and families, communities, governments and implementing organisations. Each is discussed in turn.

6.1 | Governments

Some programmes can only be sustained if they are adopted by governments because they require relatively large-scale systems and recurrent funding for their implementation. The implications for governments are obvious but no less significant because of that. For governments to be able to adopt and implement a programme, they must:

i. Prioritise the issue that the programme addresses over competing issues and rate it sufficiently highly to allocate resources to it;

TABLE 4 Learning from fidelity and contextualisation

Context	Mechanism	Outcome
High fidelity implementation of the programme in multiple sites. Coordinated and comparative realist evaluation across sites, with data about participant subgroups, mechanisms, outcomes and contextual factors. Opportunities for shared analysis of, and meaning-making from, evaluation data.	Comparative analysis of data identifies outcome patterns and the context-mechanism configurations causing them. Organisational commitments to achieving the best possible outcomes for all and belief that contextualisation contributes to effectiveness. Peer learning and problem-solving.	Redesign of elements of programme to improve effectiveness for specific contexts or subgroups. Increased attention paid to fidelity to those programme principles that contribute to effectiveness.
High fidelity implementation of the programme in multiple sites. Independent organisations (or local structures) responsible for implementation and evaluation. No structure or process for coordination or communication across sites. Different approaches to evaluation in sites. Analysis undertaken at local level.	Reduced data pool (within each site) makes identification of outcome patterns more difficult. Non-comparable data across sites makes learning from comparison of findings more difficult. No peer learning or problem solving reduces insight.	Weaker evidence base for adaptation of programme. Reduced likelihood of effective adaptations.

ii. Within that issue area, prioritise the programme over alternative strategies, which is likely to require belief in the efficacy of the intervention compared to others and congruence between the programme and the government's policy objectives;

iii. Understand the policy, resource and capacity requirements for implementation;

iv. Establish systems for the programme's implementation and administration;

v. Recruit and/or train staff to provide and manage it on an ongoing basis.

These processes all carry both direct costs and opportunity costs (that is, other opportunities that must be foregone because the resources have been invested in the programme). Further, some programmes initiated by external parties, including INGOs, have strong philosophical foundations—inclusiveness, democratic decision-making, equity and the like. Governments either need to see these underpinnings as consistent with their own objectives or tailor the programme to suit their own political and philosophical perspectives.

Either resourcing issues or politics may result in only some aspects of programmes being adopted by governments. Without good evidence about how and why the original programme worked, partial adoption may undermine the effectiveness of the programme—in which case, the programme (or parts of it) may be sustained but outcomes may not be.

6.2 | Communities

The three frameworks outlined above highlight several implications for the roles that communities can play in ensuring that positive outcomes from international development intervention are sustained when external funding ceases.

The first relates to 'community ownership' (albeit the extent to which this is spread across communities is debatable). Community members must be involved in interventions from their inception. This ensures that interventions are desired by communities, align with their values and will be appropriate to their needs. This is likely to be a prerequisite for community members to keep the intervention going when funding ceases.

Secondly, ownership requires participation. Without broad-based community participation, the capacity to sustain benefits (discussed below) is unlikely to be built. Moreover, participation can strengthen community cohesion and improve the ability of members to adopt and adapt an intervention as they learn key skills from one another. Broad-based participation raises the likelihood of changes in social norms and practices that are necessary for sustained outcomes as well as strengthening relationships and/or managing conflict and tensions amongst key stakeholders.

Ownership and participation are important in achieving a third crucial factor: capacity building. If an intervention is to be maintained, community members must acquire the knowledge and skills needed to implement, manage and adapt the intervention themselves before funding ceases. This might also require the establishment of new structures and capacity to undertake advocacy may also be important.

Finally, as Framework 2 above suggests, benefits will not be sustained unless communities have adequate resources. Resources here include financial resources to cover the recurrent costs of an intervention (such as overheads and salaries) as well as the costs of maintaining and replacing any physical infrastructure; human resources; and time. These must be sufficient to sustain both the intervention itself and the social infrastructure required to manage it.

6.3 | Individuals and families

The frameworks above imply quite different roles for individuals and families. Roles will vary within communities and within programmes. This, in turn, implies that there will be different *contributions to* sustainability by different subgroups of individuals and families and different *effects of* sustained outcomes for them.

Some people will be programme participants. Participating means that families have to prioritise the programme over all the other competing demands on their time—farming, working, parenting their children, participating in other community activities and so on. They have to commit (usually scarce) human and sometimes financial or material resources to the activities required.

Given the scarcity of such resources, this implies that there must be sufficient return on that investment to warrant its continuation—be that return in terms of human capital (new knowledge or skills), well-being (e.g., improved health), social capital (improved relationships), economic capital (improved livelihoods) or some combination of these. Only a very direct self-sustaining feedback loop (e.g., changed behaviour generates direct benefits that maintain motivation for sustained behaviour change) is likely to result in sustained outcomes once the direct supports of a programme are withdrawn. Moreover, given that outcomes of most programmes vary in nature and extent for different subgroups within the population, it is highly likely that sustained outcomes will also vary for different subgroups.

Other individuals will be volunteers, making a higher investment in the programme than beneficiaries and thus facing greater potential for both benefits and costs. For example, many programmes work through peer facilitators or peer educators. Peer educators often benefit more from peer education programmes than do participants (see, e.g., Wawrzynski et al., 2011). However, should the programme end, the educators stand to lose many things: a valued role in the community, a source of identity and purpose, social interaction and potentially social capital and whatever educational or material benefits the role attracted. Human capital developed through the role may be sustained for some time but knowledge and skill deteriorate if not used; maintaining other outcomes will depend in part on the individual and in part on the opportunity structure of the community.

Other individuals again will be in governance roles with some level of decision-making responsibility for the programme and the direct and indirect (e.g., cultural, political and environmental) costs and benefits it creates. Being a decision-maker carries its own set of risks and rewards; these include human capital and reputational risks and rewards, which may play out in employment and/or in political systems. Whether or not these benefits or harms are sustained should be one component of assessments of sustainability of outcomes.

6.4 | Implementation agencies

The agencies implementing interventions play a critical role in improving the likelihood that the outcomes of their interventions will be sustained. As noted in the 'resources' framework above, agencies provide material and financial resources as part of interventions. Choices about what to provide and how have been shown to effect sustainability of behaviours. Rogers and Coates (2015) found that the withdrawal of free resources and participation incentives at the conclusion of interventions could lead to loss of participant engagement in the behaviours required to sustain outcomes; developing and testing CMOs from their framework could illuminate how and why this occurs in different contexts.

In many cases, the presumption is that other parties will provide resources on an ongoing basis once the value of the intervention has been shown. The transition period and processes have been identified as particularly important for ensuring that intervention outcomes will be sustained into the future (Springfield Center 2014) with a gradual transition period recommended to improve sustainability. Transitions comprise both 'phase out' of resource provision by the original agency and often 'phase over' to other partners who may take on some aspects of resource provision (Rogers & Coates, 2015).

The 'phase over' of responsibility depends in large part on linkages with individuals, organisations and institutions pursued during projects. Implementation agencies need to facilitate these linkages from the earliest stages of programmes and collaborate in development of plans and capacity building for the adopting agency.

Where volunteers are expected to sustain programme activities, strategies are required for the initial transfer of responsibilities and for capacity building for future incumbents.

The third framework above indicated the importance of monitoring and evaluation. However, many useful indicators have not commonly been collected. Rogers and Coates (2015) noted the importance of monitoring and evaluation of continued resources, capacity and motivation; establishment of appropriate linkages; and transition to independent operation. Trandafili (2019, p. 5) notes that indicators such as 'group sustainment' are required to track how sustainable community mobilisation is likely to be. Overall, the frameworks suggest that making sustainable outcomes an organisational priority has implications for planning and design, implementation, monitoring and evaluation.

7 | CONCLUSION

This article has presented three frameworks derived from realist research or evaluation undertaken in international development programmes. Each addresses different aspects of sustainability, and each has implications for governments, communities and individuals, as well as for organisations implementing international development programmes. Standing back from the detail of the individual frameworks, the three together suggest a particular sequence for planning for sustainability.

Initially, the broad pathway to sustained outcomes should be identified. This includes determining whether the intervention itself needs to be sustained in order for outcomes to be sustained. It also includes determining the primary pathway (or pathways) required for outcomes to be sustained.

Once the intended pathway and model are known, resource requirements can be identified, with due attention to the different types of resources and their different sources. Likely risks to availability of different kinds of resources can be assessed, with strategies put in place to minimise risks where possible.

Then the relative importance of fidelity to existing frameworks or evidence, and adaptation to the specific context, can be assessed. For sustainability of outcomes, this is most likely to matter when sustained delivery of a programme is required.

The frameworks also suggest particular issues for investigation, and frameworks for analysis, in our planned programme of research into sustainability of outcomes in international development. As noted earlier, the 'pathways' framework hints at but does not yet identify different mechanisms contributing to sustainability of outcomes; these will need to be identified and the contextual factors that affect their operation identified. The resources' framework

suggests both potential indicators for evaluation and an analytic framework for comparisons of resourcing contributions where outcomes are, and are not, sustained. The fidelity and contextualisation framework suggests features of programme design and implementation that may be significant: whether, when and why they are should also be included in future investigations.

Two other frameworks cited in this paper were not developed through realist research and evaluation; they focus on factors related to outcome sustainability at the period when projects are transitioning to local ownership. They provide a potential starting point for realist investigation of how and why particular outcomes are generated in different transition, 'phase out' and 'phase over' contexts.

The research project on which our consortium is embarking will seek to examine these and some other vexed policy questions. What are the relationships between sustainment of outcomes and the capacities of local stakeholders to sustain interventions? How do they vary for different types of interventions, in different contexts? Over what time period is it reasonable to expect outcomes to be sustained, and how should those expectations vary for different types of interventions, in different contexts? What evidentiary support is there for the various frameworks of sustainability and their applicability across different sectors and settings? There are, however, two starting points on which we are agreed. The first is that understanding sustainability can only be satisfactorily investigated by starting from what is sustained (rather than predicting what might be). The second is that, given the complexity of sustainment, only a theory-based approach to investigation holds true promise. Realist evaluation takes account both of underlying causal processes and of the effects of context; can be applied across interventions, sectors and scales of systems; provides a rigorous basis for generalisation; and allows aggregation of findings across multiple research projects. We therefore contend that it is the most appropriate overarching framework for a programme of investigation into sustainability of outcomes.

DATA AVAILABILITY STATEMENT

There are no empirical data associated with the manuscript.

ORCID

Simon Feeny https://orcid.org/0000-0002-5955-0045
Gill Westhorp https://orcid.org/0000-0001-8451-2860
Emma Williams https://orcid.org/0000-0001-5590-7644

REFERENCES

Bishop, C. (2018). Ex post evaluation of technology diffusion in the African palm oil sector: the Caltech expeller in Cameroon, Benin, and Liberia. *World Development*, *112*, 233–243. https://doi.org/10.1016/j.worlddev.2018.08.022

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *The Quarterly Journal of Economics*, *129*(2), 697–752. https://doi.org/10.1093/qje/qjt057

Blattman, C., Fiala, N., & Martinez, S. (2018). *The Long Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program* (Vol. No. w24999). National Bureau of Economic Research.

Brundtland Report. (1987). *Our Common Future* (Vol. 14, pp. 291–294). United Nations.

Cekan, J. (2016). How to foster sustainability. *Global Policy*, *7*(2), 293–295.

Cisilino, F., Bodini, A., & Zanoli, A. (2019). Rural development programs impact on environment: An ex-post evaluation of organic farming. *Land Use Policy*, *85*, 454–462. https://doi.org/10.1016/j.landusepol.2019.04.016

Clarke, M., Feeny, S., & Donnelly, J. (2014). Water, Sanitation and Hygiene Interventions in the Pacific: Defining, Assessing and Improving 'Sustainability'. *The European Journal of Development Research 26*, 692–706. https://doi.org/10.1057/ejdr.2013.67

CRS. (2016). Participation by All: The Keys to Sustainability of a CRS Food Security Project in Niger, Catholic Relief Services.

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). Whats in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10, 49. https://doi.org/10.1186/s13012-015-0237-x

de Mel, S., McKenzie, D., & Woodruff, C. (2012). One-time transfers of cash or capital have long-lasting effects on micro-enterprises in Sri Lanka. *Science*, *335*(6071), 962–966. https://doi.org/10.1126/science.1212973

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, *48*(2), 424–455. https://doi.org/10.1257/jel.48.2.424

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

DFAT. (2021), Aid Programming Guide Glossary, Department for Foreign Affairs and Trade.

DfID. (2012), Broadening the Range of Designs and Methods for Evaluation, DfID Working Paper No.38, Department for International Development.

Dossou, J.-P., van Belle, S., & Marchal, B. (2021). Applying the realist evaluation approach to the complex process of policy implementation—The case of the user fee exemption policy for cesarean section in benin. *Frontiers in Public Health*, 9.

Douthwaite, B., Mayne, J., McDougall, C., & Paz-Ybarnegaray, R. (2017). Evaluating complex interventions: A theory-driven, realist-informed approach. *Evaluation*, *23*(3), 294–311. https://doi.org/10.1177/1356389017714382

Easterly, W. (2007). *The White Mans Burden: Why the Wests Efforts to Aid the Rest have Done So Much Ill and So Little Good*. Penguin.

Fafchamps, M., McKenzie, D., Quinn, S., & Woodruff, C. (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana. *Journal of Development Economics*, *106*, 211–226. https://doi.org/10.1016/j.jdeveco.2013.09.010

Faure, G., Blundo-Canto, G., Devaux-Spatarakis, A., le Guerroué, J. L., Mathé, S., Temple, L., Toiller, A., Triomphe, B., & Hainzelin, E. (2020). A participatory method to assess the contribution of agricultural research to societal changes in developing countries. *Research Evaluation.*, *29*(2), 158–170. https://doi.org/10.1093/reseval/rvz036

Feeny, S., Westhorp, G., Jennings, M., Clarke, M., Donohue, C., Boxelaar, L., Mackintosh, A., Navez-John, K., Nyanhanda, T., Castro, J., & Wheatley, N. (2017). *World Visions Child Sponsorship: An Evaluation of Five Area Development Programmes*. Report for World Vision International.

Fowler, A. (Ed.) (2013). *The Virtuous Spiral: A Guide to Sustainability for NGOs in International Development*. Routledge. https://doi.org/10.4324/9781315071374

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998–1001. https://doi.org/10.1126/science.1251178

GHPC. (2016). *Sustainability Framework*. Georgia Health Policy Center, Georgia State University. https://ghpc.gsu.edu/tools-frameworks/sustainability-framework/

Gómez-Lobo, A. (2020). Transit reforms in intermediate cities of Colombia: An ex-post evaluation. *Transportation Research Part A: Policy and Practice*, *132*, 349–364.

Graham, H., Bakare, A., Gray, A., Ayede, A., Qazi, S., McPake, B., Izadnegahdar, R., Duke, T., & Falade, A. (2018). Adoption of paediatric and neonatal pulse oximetry by 12 hospitals in Nigeria: A mixed-methods realist evaluation. *BMJ Global Health*, *2018*(3), e000812. https://doi.org/10.1136/bmjgh-2018-000812

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., & Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *The Lancet*, *371*(9610), 411–416. https://doi.org/10.1016/S0140-6736(08)60205-6

Hoffecker, E. (2021). Understanding inclusive innovation processes in agricultural systems: a middle-range conceptual model. *World Development*, *140*, 105382. https://doi.org/10.1016/j.worlddev.2020.105382

IED. (2020). 2020 Annual Evaluation Review: ADBs Project Level Self-Evaluation System. Independent Evaluation Department, Asian Development Bank.

IEG. (2015). *World Bank Project Performance Ratings—Codebook*. Independent Evaluation Group, World Bank.

Ishola, F., & Cekan, J. (2019). Evaluating the sustainability of health programmes: A literature review. *African Evaluation Journal*, *7*, 1. https://doi.org/10.4102/aej.v7i1.369

Kaaria, S. K., Kirkby, R. A., Delve, R. J., Njuki, J., Twinamasiko, E., & Sanginga, P. C. (2004). Enhancing innovation processes and partnerships. In *Centro Internacional de Agricultura Tropical (CIAT)*. International Center for Tropical Agriculture.

Kiptot, E., Franzel, S., Hebinck, P., & Richards, P. (2006). Sharing seed and knowledge: Farmer to farmer dissemination of agroforestry technologies in western Kenya. *Agroforestry Systems*, *68*(3), 167–179. https://doi.org/10.1007/s10457-006-9007-8

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., & Stein, A. D. (2009). The impact of improving nutrition during early childhood on education among Guatemalan adults. *The Economic Journal*, *119*(537), 734–763. https://doi.org/10.1111/j.1468-0297.2009.02220.x

Marchal, B., Dedzo, M., & Kegels, G. (2010). Turning around an ailing district hospital: A realist evaluation of strategic changes at HO Municipal Hospital (Ghana). *BMC Public Health*, *10*, 787. https://doi.org/10.1186/1471-2458-10-787

Merton, R. K. (1968). On sociological theories of the middle range. In R. K. Merton (Ed.), *Social Theory and Social Structure* (Enlarged ed., Vol. 1968, p. 1968). The Free Press.

Moore, J. E., Mascarenhas, A., Bain, J., & Straus, S. E. (2017). Developing a comprehensive definition of sustainability. *Implementation Science*, *12*, 110. https://doi.org/10.1186/s13012-017-0637-1

Moyo, D. (2009). *Dead Aid: Why Aid is Not Working and How There is a Better Way for Africa*. Farrar, Straus and Giroux.

Mukova-Moses, M. (2017). A case for the advancement of utilisation of realist evaluation in international development: An Oxfam focused illustration. *Evaluation Journal of Australasia*, *17*(2), 37–41. https://doi.org/10.1177/1035719X1701700206

OECD. (2018). *OECD Regulatory Enforcement and Inspections Toolkit*. OECD Publishing. https://doi.org/10.1787/9789264303959-en

OECD. (2020). Aid by DAC members increases in 2019 with more aid to the poorest countries. Development Assistance Committee of the OECD.

OECD. (2021). Evaluation Criteria. Development Assistance Committee of the OECD. https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

Owusu-Addo, E., Renzaho, A. M. N., & Smith, B. (2020). Developing a middle-range theory to explain how cash transfers work to tackle the social determinants of health: A realist case study. *World Development*, *130*, 104920. https://doi.org/10.1016/j.worlddev.2020.104920

Pawson, R. (2002). Evidence-based policy: The promise of realist synthesis. *Evaluation*, *8*(3), 340–358. https://doi.org/10.1177/135638902401462448

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Sage.

Punton, M., Isabel, V., Leavy, J., Michaelis, C., & Boydell, E. (2020). Reality bites: Making realist evaluation useful in the real world, CDI Practice Paper 22, Brighton: IDS. 87–100.

Ravallion, M. (2020). *Should the Randomistas (Continue to) Rule?* National Bureau of Economic Research Working Paper 27554, National Bureau of Economic Research. https://doi.org/10.3386/w27554

Riddell, R. C. (2008). *Does Foreign Aid Really Work?* Oxford University Press.

Rigol, N., Hussam, R., & Roth, B. (2017). Targeting high ability entrepreneurs using community information: Mechanism design in the field. Working paper, Harvard University.

Rogers, B. L., & Coates, J. (2015). *Sustaining Development: A Synthesis of Results from a Four-Country Study of Sustainability and Exit Strategies among Development Food Assistance Projects*. FHI 360/Food and Nutrition Technical Assistance III Project (FANTA).

Sachs, J. (2005). *The End of Poverty: Economic Possibilities for Our Time*. Penguin.

Sarriot, E., Ricca, J., Yourkavitch, J., Ryan, L., & the Sustained Health Outcomes (SHOUT) Group. (2008). *Taking the Long View: A Practical Guide to Sustainability Planning and Measurement in Community-Oriented Health Programming*. Macro International Inc.

Schell, S. F., Luke, D. A., Schooley, M. W., Elliott, M. B., Herbers, S. H., Mueller, N. B., & Bunger, A. C. (2013). Public health program capacity for sustainability: a new framework. *Implementation Science*, *8*, 15. https://doi.org/10.1186/1748-5908-8-15

Stein, A. D., Wang, M., DiGirolamo, A., Grajeda, R., Ramakrishnan, U., Ramirez-Zea, M., Yoint, K., & Martorell, R. (2008). Nutritional supplementation in early childhood, schooling, and intellectual functioning in adulthood: a prospective study in Guatemala. *Archives of Pediatrics & Adolescent Medicine*, *162*(7), 612–618. https://doi.org/10.1001/archpedi.162.7.612

Taylor, B. (2014). Who wants to give forever? Giving meaning to the term sustainability in international development. *Journal of International Development*, *26*, 1181–1196. https://doi.org/10.1002/jid.3033

The Springfield Centre. (2014). *The Operational Guide for the Making Markets Work for the Poor (M4P) Approach* (2nd ed.). funded by SDC & DFID.

Tierney, M., Nielson, D., Hawkins, D., Roberts, J., Findley, M., Powers, R., Parks, B., Wilson, S., & Hicks, R. (2011). More dollars than sense: Refining our knowledge of development finance using AidData. *World Development*, *39*(11), 1891–1906. https://doi.org/10.1016/j.worlddev.2011.07.029

Tilley, N., & Westhorp, G. (2020). Evaluation research, quantitative. In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE Research Methods Foundations*. SAGE Publishing. https://doi.org/10.4135/9781526421036872988

Trandafili, H. (2019). *World Vision Learning Brief: What Does Sustainability Look Like Post-Program?* World Vision US.

Uphoff, N., & Wijayaratna, C. M. (2000). Demonstrated Benefits from Social Capital: The Productivity of Farmer Organizations in Gal Oya, Sri Lanka. *World Development*, *28*(11), 875–1890. https://doi.org/10.1016/S0305-750X(00)00063-2

USAID. (2020). *What Does it Take to Sustain Water, Sanitation and Hygiene Outcomes? Lessons from Six Ex-Post Evaluations*. United States Agency for International Development.

Walker, S. P., Chang, S. M., Vera-Hernández, M., & Grantham-McGregor, S. (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics*, *127*(5), 849–857. https://doi.org/10.1542/peds.2010-2231

Wawrzynski, M. R., LoConte, C. L., & Straer, E. J. (2011). Learning outcomes for peer educators: The national survey on peer education. *New Directions for Student Services*, *2011*(133), 17–27. https://doi.org/10.1002/ss.381

Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation*, *18*(4), 405–420.

Westhorp, G. (2013). Developing complexity consistent theory in a realist investigation. *Evaluation*, *19*(4), 364–382. https://doi.org/10.1177/1356389013505042

Westhorp, G. (2014). *Realist Impact Evaluation: An Introduction*. ODI Methods Lab Publication, Overseas Development Institute. https://cdn.odi.org/media/documents/9138.pdf

Westhorp, G. (2018). In N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan, & S. Dalkin (Eds.), Understanding mechanisms in realist evaluation and research, *Doing Realist Research*. Sage. https://doi.org/10.4135/9781526451729.n4

Westhorp, G., Williams, E., Booth, V., & Ball, D. (2017). *Are Our Programs Working?* Charles Darwin University final report for World Vision UK.

World Bank. (2020). *IEG World Bank Project Performance Ratings*. World Bank. https://finances.worldbank.org/Other/IEG-World-Bank-Project-Performance-Ratings/rq9d-pctf

WVI. (2016). *World Visions Drivers of Sustainability*. World Vision International. https://www.wvi.org/sites/default/files/WV%27s%20Sustainability%20Drivers%20-%20Summary%20-%202-15.pdf

---

How to cite this article: Feeny, S., Westhorp, G., & Williams, E. (2023). Understanding sustainable outcomes in international development: Towards a realist evaluation framework. *Journal of International Development*, *35*(1), 21–42. https://doi.org/10.1002/jid.3662
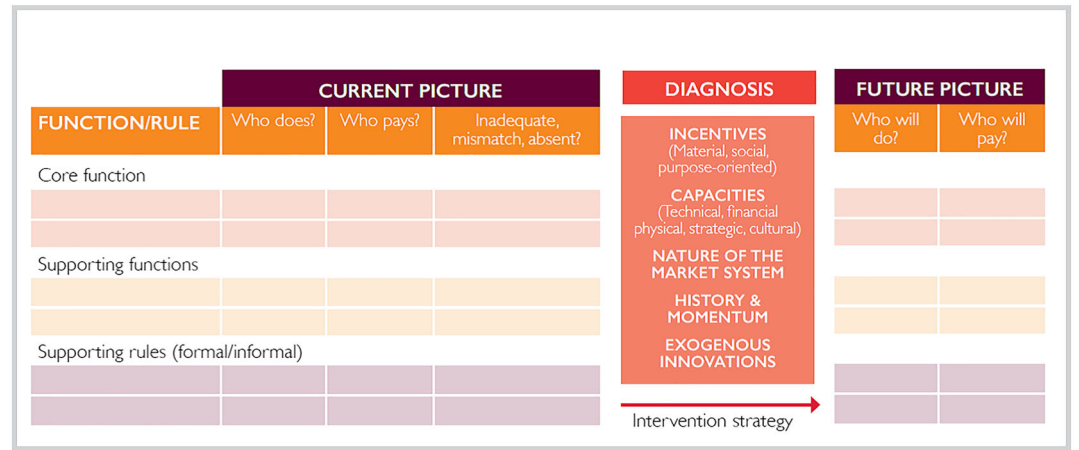

APPENDIX A

FIGURE A1 Sustainability analysis framework. *Source*: The Springfield Centre (2014)

FIGURE A2 Key questions and indicators of sustainability and scale. *Source:* The Springfield Centre (2014)

Independent investment: Former (ie pilot phase) partners have invested in upholding, or qualitatively/quantitatively improving upon, the change(s) adopted, without programme support

Target group benefits are sustained: The programme's beneficiary groups continue to benefit after programme support to the pilot-phase partner ends

System responsiveness and receptiveness: Non-competing players respond with practice changes of their own that enable early and late 'adopters' to operate more effectively or further improve their performance. Often borne out in: new services available, new responsibilities assumed, streamlined procedures, amended regulations/legislation, etc

Ability of 'adopters' to cope with shocks: Evidence that the practice changes of early and late 'adopters' can withstand adverse events or negative reactions from other market players: eg in response to weather extremes, commodity price shocks, moderate political pressures etc

ADAPT
If you left now, would partners build upon the changes they have adopted, without us?

RESPOND
If you left now, would the system be supportive of the changes introduced (allowing them to be upheld, grow, evolve)?

ADOPT
If you left now, would partners return to their previous way of working?

EXPAND
If you left now, would target group benefits depend on too few people, firms, or organisations?

Partner contribution: Partner invests financially and non-financially in piloting a new way of working. Significant contributions from a pilot-phase partner demonstrate the value they have attached to what has been trialled, commitment, and therefore a degree of ownership

Long-term viability / benefit of practice changes: The nature of this indicator depends upon the commercial-mindedness of the partner in question. For more commercially-motivated partners: evidence of the commercial viability of practice changes, improvements to competitiveness, or positioning to realise longer-term, non-immediate benefits. For less commercially-motivated partners: viability of practice changes, fulfilment of priorities, etc

Partner satisfaction and intent to continue: Partner's satisfaction with results from the pilot, interest in and ownership over learning that emerges from the pilot, partner willingness to assume all recurrent costs by pilot end, concrete plans to invest in upholding practice changes without the programme

Partner ability to continue: Partner possesses the organisational capabilities and human resources (eg financial, technical, etc) to uphold the practice changes piloted

Target group's satisfaction and benefit: Beneficiary groups are satisfied with and benefiting from the outcomes of partner behaviour / practice changes

Competitors or similar types of organisations crowd-in: The nature of this indicator depends upon whether or not the pilot intervention engaged with commercially-motivated players. For more commercially-motivated initiatives: competitive responses from similar players, positive outcomes from engagement with 'scale agents' or 'apex' players, etc. For less commercially-motivated initiatives: signs of influence and 'spread' inspired by the initial innovation

Ability to accommodate competition or collaboration (depends on the nature of the system): For more commercially-motivated initiatives: reduced barriers to entry faced by competitors and new players (ie late adopters). For less commercially-motivated initiatives or more collaborative systems: leadership and effectiveness of representative organisations, undertaking of joint ventures, respect for rules/regulations/standards, including voluntary codes, etc
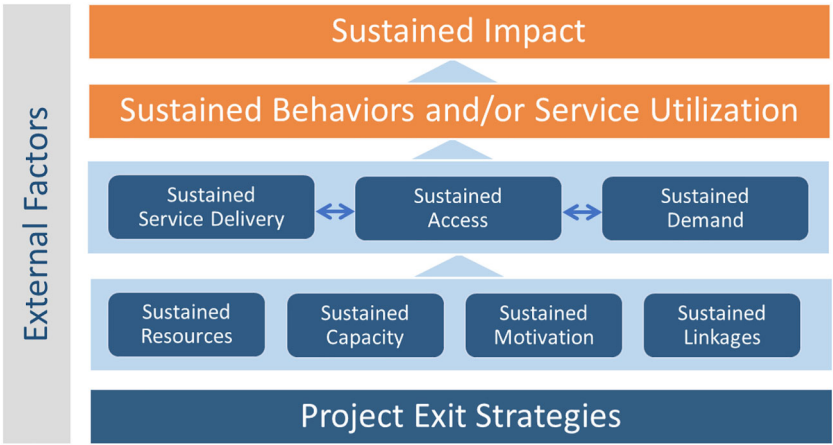
FIGURE A3 Sustainability and exit strategies conceptual framework. *Source*: Rogers and Coates (2015). Adapted from Coates and Kegode. 2012. 'Kenya Exit Strategies Study Round 2 Report'. Unpublished, submitted to FANTA April 8