

IDS Bulletin

Transforming Development Knowledge

Volume 49 | Number 4 | September 2018

THE MILLENNIUM VILLAGES: LESSONS ON EVALUATING INTEGRATED RURAL DEVELOPMENT

Editor **Chris Barnett**



Notes on Contributors	iii
Foreword	
Richard Longhurst	vii
<u>Introduction: Lessons from the Millennium Villages Evaluation; Where Next for Integrated Development?</u>	
Chris Barnett	1
Integrated Development, Past and Present	
Edoardo Masset	17
The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies	
Edoardo Masset, Giulia Mascagni, Arnab Acharya, Eva-Maria Egger and Amrita Saha	33
Assessing Value for Money in Integrated Development Programmes – The Case of a Millennium Villages Project in Northern Ghana	
Arnab Acharya and Tom Hilton	53
Abductive Reasoning to Explain Integrated Development: Lessons from the Multi-Method Evaluation of the Millennium Villages Project	
Dee Jupp and Chris Barnett	67
Can Immersion Research Add Value in Understanding Integrated Programme Interventions?	
Dee Jupp, David Korboe and Tony Dogbe	83
Learning About Integrated Development Using Longitudinal Mixed Methods Programme Evaluation	
Emily Namey, Lisa C. Laumann and Annette N. Brown	97
Applying Factorial Designs to Disentangle the Effects of Integrated Development	
Holly M. Burke, Mario Chen and Annette N. Brown	115
Glossary	129

Introduction: Lessons from the Millennium Villages Evaluation; Where Next for Integrated Development?

Chris Barnett¹

Abstract This *IDS Bulletin* explores recent evidence on integrated approaches in rural development. Since the 1970s, holistic and multisectoral programming has been based on the synergistic potential of achieving more through a 'big push' of doing lots together. The Sustainable Development Goals (SDGs) also suggest a greater need to understand and address interconnectedness across different sectors. Drawing extensively on a recent impact evaluation of the Millennium Villages Project in northern Ghana, this issue presents a series of articles on the challenges of evaluating integrated development. The articles explore the challenge of assessing synergy; the cost-effectiveness of integration; the value of mixing methods; and dealing with multiple outcomes on different timelines. This introduction concludes by suggesting a narrower testing of combinations of interventions in different contexts, and incrementally building the evidence base; rather than 'doing everything together', where a lack of impact combined with weak mid-range theory can limit learning about what works and why.

Keywords: integrated development, integrated rural development, Millennium Villages, impact evaluation, mixed methods, multisector, complexity, Ghana.

1 Background and context

Holistic, multisectoral programming – working in ways that consider the interrelated aspects of people's lives – is intuitively appealing. Marginalised or poor people are unlikely to consider life as a series of disconnected sectoral issues or challenges. As such, introducing interventions across several sectors at once should (in theory at least) be mutually beneficial: increased household incomes through improved farming practices may reduce the need for child labour and thereby increase school attendance. More children attending school may also benefit from reduced teacher absenteeism due to fewer sick days

© 2018 The Author. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.158



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non Commercial No Derivatives 4.0 International licence (CC BY-NC-ND), which permits use and distribution in any medium, provided the original authors and source are credited, the work is not used for commercial purposes, and no modifications or adaptations are made. <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK
This article is part of *IDS Bulletin* Vol. 49 No. 4 September 2018 'The Millennium Villages: Lessons on Evaluating Integrated Rural Development'.

because of better health care. Improved health care may similarly result in household members having more time to cultivate the land, thus relying less on child labour. And so on.

Meanwhile, there is a renewed interest in interconnectedness, including in finding ways that work across different sectors to address the Sustainable Development Goals (SDGs). Inherent in the SDGs is the recognition that the goals are interdependent, and often the key to success in achieving one goal will involve tackling issues more commonly associated with another. In this way, it is not enough to simply monitor the SDG indicators; countries need to know (and evaluate) how policies and programmes will *together* be effective (Jimenez and Puri 2017). This *IDS Bulletin* goes on to argue that for this reason there are aspects of integrated development that continue to remain relevant to achieving the SDGs. Indeed, the SDGs are more integrated than the previous Millennium Development Goals (MDGs), and although the specific goals are similarly unconnected, they set a global framework for integration (Schwandt *et al.* 2016).² As Le Blanc points out:

In designing and monitoring their work, agencies concerned with a specific goal (e.g. education, health, economic growth) will have to take into account targets that refer to other goals, which, due to the normative clout of the SDGs for development work coming forward, may provide stronger incentives than in the past for cross-sector, integrated work (2015: 9).

Given this policy agenda, it is therefore intriguing that integrated development projects have largely fallen out of favour since their heyday in the 1970s and 1980s (as Masset charts in this issue of the *IDS Bulletin*) – although there is more recent renewed interest in certain quarters.³ Indeed, in writing this *IDS Bulletin*, it felt at times like we were ‘swimming against the tide’ of development. Of course, this is understandable: firstly, integrated, multisectoral development has a close association with the unwieldy integrated rural development (IRD) projects of the 1970s and 1980s, which have largely been discredited (as Masset, this *IDS Bulletin*, describes). And secondly, one of the most high-profile integrated projects of recent years, the Millennium Villages Project (MVP) is framed within the pre-2015 MDG era; and itself has been the subject of much debate and critique (e.g. Clemens and Demombynes 2013; Wanjala and Muradian 2013). Integrated development projects seem rather passé, replaced by a growing interest in understanding complexity, systems-thinking, and managing adaptively – a topic we return to at the end of this article.

Perhaps, though, there is another reason why ‘integrated development’ approaches have yet to resonate fully with the SDG agenda. The new demands of the SDG agenda shift the emphasis from a more sectoral/MDG focus to one that recognises and capitalises on interconnections needed to drive sustainable development. In this way, integrated or holistic programming moves beyond a purely cross-sectoral approach (‘health

and education', 'agriculture and governance', and so on) and raises new challenges around aligning economic growth more broadly with social inclusion and long-term environmental sustainability (Barnett and Eager 2017; Thomas 2017). Understanding the progress towards sustainable development requires greater consideration of the interrelations between SDGs and how synergistic actions in one area may drive increased benefits in another (Wiesen and Prokop 2015). Whereas earlier IRD projects added social interventions (health and education) to traditional agricultural approaches, the SDGs go further in the need to link all three forms of capital: physical, social, and natural. For this reason, linear, cause/effect approaches to evaluation are often seen as insufficient to understand such complex interactions and the contextual variation that is influencing progress towards the SDGs (Befani, Ramalingam and Stern 2015; Picciotto 2015).

One of the observations in writing this *IDS Bulletin*, however, is how much the evidence base has improved, with a now greater understanding of where knowledge gaps exist (including the ones highlighted in this issue). This provides a sense of optimism. As Masset (this *IDS Bulletin*) points out, we have come a long way in terms of monitoring and evaluation in the past 50 years. Citing the Operations Evaluation Department of the World Bank,⁴ Masset discusses older IRD projects from the 1970s and 1980s thus: 'As for the evidence in support of the poor impact of IRD, this is totally non-existent. The evaluation methods used at the time [1980s] have been described as underdeveloped and intuition was often used instead of data.' Since then, Masset (this *IDS Bulletin*) goes on to explain that the quality of evaluation has significantly progressed, with recent incarnations of integrated development often being evaluated independently, and in rigorous ways – using mixed method, experimental, and quasi-experimental designs. Several such examples are drawn upon extensively in this issue, including the evaluation of the MVP in Ghana (Acharya and Hilton; Jupp and Barnett; Jupp, Korboe and Dogbe, all this *IDS Bulletin*), and of projects in Uganda (Namey, Laumann and Brown, this *IDS Bulletin*) and South Africa (Burke, Chen and Brown, this *IDS Bulletin*); as well as a recent systematic review of the evidence for integrated projects (Ahner-McHaffie *et al.* 2018), plus an evidence map of over 500 impact evaluations applied to integrated, multisector approaches.⁵

This accumulation of evidence still presents us with a central challenge, however: it shows that integrated development projects can be effective (with long-lasting effects on poverty reduction), but we do not know enough about whether any increased effect is *due to* integration – or whether it would be better achieved through separate components (Ahner-McHaffie *et al.* 2018; Masset, this *IDS Bulletin*; Masset *et al.*, this *IDS Bulletin*). Indeed, most experimental evaluations of integrated development programmes compare the integrated project with a 'without' group and therefore do not assess the impact from synergy *per se* (Ahner-McHaffie *et al.* 2018); in contrast to the study described by Burke *et al.* (this *IDS Bulletin*), which

compares the 'with' integrated intervention against both a 'without' group and 'with' the single (unintegrated) interventions.

We therefore use this *IDS Bulletin* to reflect on where we are with integrated development, with a particular (though not exclusive) focus on lessons from evaluating the Millennium Villages Project in northern Ghana. Why such a focus? In part, this is because the MVP has been one of the most prominent examples of integrated development in recent years with new evidence now emerging about its effectiveness (Mitchell *et al.* 2018; Barnett *et al.* 2018). Perhaps more importantly though, it is because the MVP is linked to a prominent international agenda to demonstrate how the MDGs could be achieved at a local, village level (Sachs 2018; Mitchell *et al.* 2018). In this way, it is timely to reflect on the MVP's effectiveness, and its relevance for learning about localising the SDGs (which themselves are even more interconnected than the previous MDGs).

In the remainder of this article, we explain more about the MVP, discuss each of the articles, and draw out a number of emerging themes. The aim is to draw together a diverse set of perspectives, and through this rich methodological pluralism, to step back from the question of whether particular integrated projects are effective (or not): and instead consider what we can learn from such approaches (around complementarities, synergies, etc.) that might be useful to take forward into a research and development agenda for the SDGs.

1.1 The Millennium Villages Project

Following on from the United Nations (UN) Millennium Summit in September 2000, world leaders committed their nations to adopting the Millennium Declaration, and a set of targets to reduce extreme poverty by 2015 (known as the MDGs). In 2002, the Millennium Project (not to be confused with the subsequent Millennium Villages Project) was commissioned by the late UN Secretary-General, Kofi Annan, and directed by Professor Jeffrey Sachs, to create a global plan of action to achieve the MDGs (UN Millennium Project 2005; Sachs and McArthur 2005). Drawing on the Millennium Project's findings, the MVP was initiated in 2005 to show how the MDGs could be achieved using an integrated and scaled-up set of targeted investments covering everything from food production and nutrition to education, health services, roads, energy, communications, and water supply.

Over the next ten years, the MVP approach was implemented in 14 rural sites across ten African countries and covering all the main agro-ecological zones. After ten years, a retrospective, endline evaluation across ten countries showed statistically significant impact for 30 of the 40 outcomes, with on average around one third of its targets being reached (Mitchell *et al.* 2018). The impact evaluation of the MVP in northern Ghana – understood to be the only independent evaluation of the MVP – shows a statistically significant impact on seven out of the 28 MDG outcome indicators, and raises concerns about such

integrated development as a cost-effective model (Barnett *et al.* 2018). In this *IDS Bulletin*, we take an appreciative look at what we can learn from the evaluation about integration, and make links to other studies of integrated development (of projects in South Africa and Uganda).

2 Emerging themes

As Namey *et al.* (this *IDS Bulletin*) and indeed several other authors highlight in this issue, integrated projects are by their nature often complex and messy – and this throws up new challenges for researchers and evaluators attempting to assess the benefits of simultaneous implementation. However, such methodological and conceptual challenges also risk opening up a Pandora’s box of ideas and approaches that would be beyond the scope of just one *IDS Bulletin*. As such, the focus here is more narrowly on a strand of thinking that originates from the early IRD projects right through to the latest MVP (as Masset, this *IDS Bulletin*, describes). This means that for the purposes of this *IDS Bulletin*, we define integrated development more narrowly in terms of the simultaneous implementation of sectoral activities to achieve (through interconnection and synergy) a greater overall impact than if done separately.⁶ For example, by combining health with agriculture, or education with economic development. This is different to a more expansive (systems-based) perspective of achieving ‘more than the sum of its parts’ through integration; where different interventions aim to work together to nudge key tipping points in the wider system and achieve transformational change (i.e. beyond immediate outcomes). In contrast to the main focus of this *IDS Bulletin*, the shift implied here is instead at a higher systems-level, such as towards a pro-poor market system or a less carbon-orientated economy.⁷

The articles in this *IDS Bulletin* are deliberately diverse (covering everything from systematic reviews to randomised trials, to mixed method designs and immersion approaches) to accommodate a wide range of perspectives. Nevertheless, when reviewed together, they highlight several reoccurring themes worth discussing in more detail, including: how to better understand synergy; the costs of integration; the added value of mixing methods for integrated projects; and the challenge of dealing with multiple outcomes on different timescales. Each is considered in turn.

2.1 Theme 1: the challenge of assessing synergy effects

Theoretically, introducing interventions in several sectors should be mutually beneficial, with the perceived advantage of integrated approaches creating synergy from ‘doing more together’ than if it was done separately. The challenge is, as Masset (this *IDS Bulletin*) points out, that the mechanisms of synergy are little understood. Indeed, synergy is something that projects struggle to capture in several different ways, including: (1) *in theory*, as Jupp and Barnett (this *IDS Bulletin*) describe in their ‘hunt’ for the MVP theory of change; and, as do Namey *et al.* (this *IDS Bulletin*) when faced with tens of thousands of possible combinations for the pathways to change for integrating economic and family strengthening; (2) *in evaluation design*, as Namey *et al.* (this *IDS Bulletin*)

found in formative work necessary to better understand the linkages between just two different sectoral projects, but also Jupp and Barnett (this *IDS Bulletin*) in the contestation of evidence and perspectives to reach a more thorough understanding of MVPs' theory of change; and, (3) *in cost-effectiveness analysis* (as both Masset *et al.* and Acharya and Hilton (both this *IDS Bulletin*) testify in their articles). The tendency seems to be for synergy to be dealt with either implicitly (as a series of assumptions that are not tested), or through reference to some grand theory that relies on complementarities (such as the 'poverty trap' in the MVP's case). As a recent systematic review surmises:

We know from the high number of randomized evaluations included here that report positive findings that in many contexts integrated, multi-sector interventions have produced impact... very few impact evaluations to date were designed to specifically examine the synergistic and interaction effects that are potentially associated with integrated programming (Ahner-McHaffie *et al.* 2018: 10–11).

There are several possible ways to address this challenge. Masset (this *IDS Bulletin*), for example, argues that what is needed is a middle-level theory: 'a theory of integrated rural development, of how it was supposed to work and under what conditions, was never explicitly elaborated and it is still absent today'. Although as Jupp and Barnett (this *IDS Bulletin*) point out, such theories are difficult to produce *ex ante* for integrated projects. Using the MVP example in Ghana, they argue that due to the complex nature of multisectoral work, there is a need to develop theory abductively in order to enrich and counterbalance the tendency towards more narrow theories based on the (privileged) perspectives of project staff or evaluators.

Burke *et al.* (this *IDS Bulletin*) goes on to helpfully untangle the synergy challenge: they argue that although the synergy question is typically stated as ' $1 + 1 > 2$ ', this can become complicated if integrated programmes do not achieve the full sum of the single intervention effects. Instead, a less stringent test of ' $1 + 1 > 1$ ' could still be considered valuable if integrated projects are shown to produce more than can be attained with a single intervention. Using a randomised control study, with different treatment arms they test combinations of the two interventions: a combination of economic strengthening and HIV prevention education (with both); HIV prevention education alone (with one); economic strengthening interventions only (with the other); and a control of no additional interventions (without).

Of course, in many cases, randomised control studies like this are not feasible. A randomised design was initially suggested to estimate the impact of the MVP (Clemens and Demombynes 2011) but later considered to be barely feasible (and not necessarily desirable), mainly because the MVP is implemented in a cluster of communities – with the cluster being the unit that should be randomised (Masset *et al.* 2013). And, even in simpler forms of integration than the MVP, such as the

integration between two interventions, a randomised design may not be possible – as Namey *et al.* (this *IDS Bulletin*) describe for two projects in Uganda that aim to integrate family and economic strengthening. In such cases, there is a role for observational and formative studies to explore integration, build the evidence base, and potentially provide the focus for subsequent evaluations that ‘test’ synergy.

2.2 Theme 2: what is the cost of integration, and is it worth it?

A second common theme highlighted in this *IDS Bulletin* is the cost of integration. As Masset (this *IDS Bulletin*) points out, the old IRD projects of the 1970s and 1980s relied on centralised coordination units, often implemented in parallel to existing institutional structures in deprived areas. More recent integrated projects have tended to rely on implementing units that are more flexible and better designed. Still, we do not know enough about the cost of integration and importantly, whether these extra costs are outweighed by the additional (synergy) benefits. Given this lack of knowledge, the main focus of the articles in this *IDS Bulletin* are on how we might begin to know more.

To this end, Masset *et al.* (this *IDS Bulletin*) undertook a systematic search of studies that assess cost-effectiveness and find a paucity of such studies. In many ways, this is unsurprising. Cost-effectiveness analysis (CEA) has tended to focus on isolating specific policy options (e.g. antimalarials, deworming tablets, bed nets, etc.) where costs and effects can be more easily attributed. Plus, as Thomas (2017) points out, CEA has been on the decline especially in the multilateral development banks, partly because of the difficulty in applying such techniques in the social sectors. Indeed, it is notable that the growing interest in complex, multisectoral interventions over recent years has not (to date) been accompanied by a similar rise in CEAs: with Masset *et al.* (this *IDS Bulletin*) finding only seven studies in the literature that address the CEA of complex projects (i.e. those with several interacting activities leading to multiple outcomes); and none of the studies reviewed address the presence of synergy effects. They conclude:

[O]ur review found... no widely applicable methodologies, and a number of practical problems in measuring the costs and effects of [integrated, complex] interventions... what appears to be more urgently needed is the discovery of methodologies able to aggregate outcomes and disaggregate costs, and a more systematic approach to cost-effectiveness of complex interventions.

As Namey *et al.* (this *IDS Bulletin*) emphasise, however, there is an often very practical challenge of gathering sufficiently disaggregated cost information for such interventions, but there are other challenges that are more unique to integration (Masset *et al.*, this *IDS Bulletin*): (1) multiple outcomes are not easily aggregated into a single index of effectiveness, such as an overarching welfare indicator; and (2) project budgets cannot be easily disentangled between different project activities and assigned to intended outcomes. The former is especially challenging for integrated projects (as discussed further below) because multiple

outcomes may be operating to different timelines and measured in very different ways – as Burke *et al.* (this *IDS Bulletin*) discuss in their evaluation of economic strengthening (through self-reported data) alongside HIV risk behaviours (through biological specimens).

From the literature, Masset *et al.* (this *IDS Bulletin*) suggest four possible approaches to address this challenge of assessing the cost-effectiveness of integrated projects: cost–consequence analysis, cost–apportionment, cost–utility analysis, and cost–benefit analysis. While each has potential, the reviewed studies share some limitations, including not considering all intended and unintended outcomes, not reporting confidence intervals to inform policy conclusions, and practical difficulties obtaining cost data. Acharya and Hilton (this *IDS Bulletin*) go on to discuss the example of cost–consequence analysis applied to the MVP in northern Ghana. The authors hypothesise that for outcomes in health, education, and income at least, a synergy effect of the MVP would be detected if the project's benefits relative to costs are greater than comparable single sector interventions. This is a similar argument to Burke *et al.* (this *IDS Bulletin*) that ' $1 + 1 > 1$ '.

While Acharya *et al.* (this *IDS Bulletin*) find that in the northern Ghana MVP case the benefits do not outweigh the costs of doing it separately, they caution against jumping to a hasty policy conclusion. Instead, the article argues that the interpretation of CEA for integrated projects is almost as important as the calculation itself. This, they argue, is because of a number of features of integrated projects: first, the need to contextualise the costs, given that large-scale integrated projects (such as the MVP) are effectively compensating for a minimum level of local government activities in resource-poor regions, and this may require greater funding than elsewhere. Second, some integrated projects are also attempting system-level changes (such as building up the health or education system through multiple interventions). This makes them difficult to compare to standalone projects that may have quicker, short-term effects – as a consequence of indeed benefiting from the system-level investments of others. Third, in part due to the reasons cited above, it may be that synergy is only observable over a much longer period rather than a typical project lifespan.

Given the current state of knowledge, it is difficult to conclude much on the costs of integration, except that this is a field that is underdeveloped, and it remains to be seen how much policy demand there is for fully understanding the cost-effectiveness of integrated projects – although Thomas (2017) advocates that it still has an important role in SDG-related evaluation.

2.3 Theme 3: the value of mixing methods

Jupp and Barnett, Jupp *et al.*, and Namey *et al.* (all this *IDS Bulletin*) highlight the value in deploying mixed method approaches when evaluating integrated projects. Of course, mixed methods are routinely used in development evaluation, and there are many ways by which to combine methods. However, what comes across strongly in this collection of articles is that the inherent characteristics of integrated projects, such as

synergy effects, and the fuzziness that surrounds it, place a greater burden on understanding the phenomenon from different research traditions. Indeed, this is a challenge highlighted by Jimenez and Puri in reviewing the future of impact evaluations in respect of the 2030 SDG agenda:

Complexity poses a substantial challenge to impact evaluations. Many programs involve a multitude of sectors: for example, livelihood programs include interventions in water provision, sanitation, income-generation activities, and health. This usually means that causal pathways are not direct, are crosslinked, and are nonlinear. Separately, it also means that there are a multitude of sectors that every program is aiming to target (2017: 354).

Furthermore, as both Jupp and Barnett and Jupp *et al.* (this *IDS Bulletin*) point out, this is not only a matter of fully *understanding* how integration is expected to work (such as using theories of change), but also *how people respond and in turn, change the intervention*.

Jupp and Barnett (this *IDS Bulletin*), for instance, argue that we need to go beyond the traditional approaches to ‘mixing’ that rely on notions of triangulation, careful integration, and the sequencing of methods. Using the MVP evaluation as an example, the article highlights how emergent theory leads to a deeper understanding that better captures how local people experience change. The authors explain that ‘the drive for theory-based evaluations – where evidence is primarily collected against the theory – can limit the ability to see beyond mostly [self-]confirmatory explanations of how change occurs’. The article describes how the team deliberately pitted the *ex ante* and *ex post* theories against each other, and goes on to explain that: ‘This requires contesting different theoretical lenses to the data with the intention of producing more thoughtful analysis, and leading to abductive (best fit) explanations; which in the end would be more useful to understanding the theory of change behind the impact of the MVP’.

For example, the initial theory of change suggested that the project’s activities (fertiliser, seeds, tractor hire, etc.) had been directly responsible for improvements in agricultural productivity. However, through the contestation of different perspectives and evidence, the eventual ‘best fit’ explanation highlighted other factors that contributed to the impact. These included: the change from growing millet to maize, for which the MVP could take some credit, and which enabled farmers to plant a second crop comprising of cowpeas – a crop not promoted by the project; the subsequent growing of cowpeas which was supported by the sharing of knowledge between farmers; and the influx of dealers supplying chemicals – also non-project activities.

Jupp *et al.* (this *IDS Bulletin*) further highlight the importance of understanding people’s experience of integrated projects, because their perspective can diverge significantly from those of project staff (or evaluators). Their article focuses on immersion studies that go beyond

participatory processes undertaken on *our* own terms in *our* spaces, and instead argue for an engagement with *their* spaces and *their* daily life.

While Namey *et al.* (this *IDS Bulletin*) do not undertake the same level of engagement, the authors show how open-ended interviews provide similar insights into participant-observed effects and what it means to live their lives; even to the extent of enquiring about how local people perceive the synergy between family and economic strengthening activities. Rich descriptions from household-level case histories were also used to gather both caregiver and child perspectives on complementary and synergy effects.

Overall, in terms of mixing methods, these articles highlight how a range of qualitative methods help untangle the interactions between different activities, begin to surface unintended or unexpected consequences of integration, and build the evidence base for future projects or areas of enquiry for integrated development. While Burke *et al.* (this *IDS Bulletin*) highlight a situation where the intervention can be altered (designed) and randomised to provide a more clear-cut 'test' of synergy, it is perhaps in the majority of cases that this is not possible.

2.4 Theme 4: multiple outcomes, multiple timelines

One final challenge of integrated projects that this *IDS Bulletin* highlights is that by attempting to address more than one outcome at the same time, each may in fact be operating in a different way and to different timelines of change. Acharya *et al.* (this *IDS Bulletin*) highlights this at the systems-level, where a project like the MVP covers most of the functions of local government (health services, education, agricultural extension, income-generating activities, farmer inputs, road building, etc.). In a resource-poor context, such as northern Ghana, the timeline for realising systems-level benefits may be beyond the project's lifespan. In contrast, standalone (simpler) projects may benefit from the investment of others (by government or development projects), especially where there are otherwise under-resourced systems for health and education. As such, a comparison of synergy effects between the two may underestimate the former – simply because the integrated project is addressing systems-level changes that are not realised as quickly as the more specific, direct outcomes of the standalone project.

But even within the same integrated project, different outcomes may be operating in different ways. For example, Burke *et al.* (this *IDS Bulletin*) show how HIV risk prevention and economic strengthening operate on different timelines: most youth will take a long time to save enough money before they become financially independent enough (via education, skills training, setting up a business, etc.) so that they no longer need to engage in transactional or intergenerational sex to meet their needs. In contrast, participating in the HIV prevention intervention, youth are more likely to engage in protective behaviours, which reduce the HIV risk outcomes more quickly. As such outcomes develop differently in time, Burke *et al.* (this *IDS Bulletin*) overcame this challenge by collecting the endline data twice (once at project end, and

once some time afterwards). This allows for an exploration of how the different outcomes of integration are realised at different times.

3 Concluding remarks

In recent years, there has been a growing appreciation of complexity, including how interventions are situated in a set of complex dynamic interactions within a broader (natural or social) system. Integrated projects such as the MVP are inherently complex, as pointed out in this issue by several authors. Masset (this *IDS Bulletin*), for example, highlights how the complexity of multi-input and multi-output interactions (and the consequent synergistic effect) is little understood and poorly theorised, at least for large-scale, integrated development projects. Masset *et al.* and Acharya *et al.* (both this *IDS Bulletin*) also show the methodological gaps in assessing and interpreting the costs and benefits of combining multiple interventions; while Jupp and Barnett, and Namey *et al.* (both this *IDS Bulletin*) show how difficult it is in reality to theorise and understand these multidimensional interactions; with Jupp *et al.* (this *IDS Bulletin*) demonstrating an added layer of complexity when the perspectives of local people are more fully understood; and finally, Burke *et al.* (this *IDS Bulletin*) show that even in a context where a technically clearer test of synergy was possible, there were still significant methodological challenges (e.g. with different outcomes needing to be measured in different ways, more than one endline to capture how multiple outcomes change over time, etc.).

Yet, despite all these challenges, we have the toolbox of designs and methods to evaluate integrated projects – and are better placed to do so than ever before. Rather, one of the lessons from the evaluation of the MVP in northern Ghana in particular, is that the ‘complexity’ label masks a lack of conceptual clarity that has implications for both project design and evaluation. Most integrated projects – by doing many things at once to achieve multiple objectives – are attempting to achieve ‘more together than apart’ in two different ways: firstly, through synergy to achieve improved outcomes ($‘1 + 1 > 2’$, although as Burke *et al.* (this *IDS Bulletin*) points out, more realistically $‘1 + 1 > 1’$); and secondly, a shift beyond the project to affect the wider system (such as ‘breaking the poverty trap’, in the case of the MVP, a sort of $‘1 + 1 = X’$).

Each has implications for the way in which we design more evaluable interventions (e.g. by project implementation staff), as well as ‘nested’ evaluation designs (e.g. by commissioners and evaluators).⁸ Focusing primarily then on the first challenge in this *IDS Bulletin* (assessing enhanced outcomes achieved through synergy), there are a number of implications from the articles:

- 1 Developing more specific (i.e. empirically testable) mid-range theories about how different activities and interventions are expected to interact is key (Masset, this *IDS Bulletin*). One of the challenges of the MVP evaluation was that once evidence of impact was found to be less than expected, it was then very hard for project staff and evaluators to

- untangle and explain why this might be so (Barnett *et al.* 2018), and thus suggest improvements. As Jupp and Barnett (this *IDS Bulletin*) point out, using abductive reasoning can also help develop such theories over time.
- 2 To achieve this, as Namey *et al.* and Burke *et al.* (both this *IDS Bulletin*) skilfully demonstrate, the key is focusing on narrower combinations where (say) just two interventions are being integrated.⁹ While still challenging, this gives an opportunity to better specify and robustly test synergy, as well as producing evidence that is more likely to have wider applicability (i.e. lessons on combining family strengthening and economic development has relevance to those in the sector, whereas large-scale multisector projects are less likely to be replicated).
 - 3 Where possible, robustly testing different combinations (with *integration*, with *single* interventions only, *without*). Where this is not possible, to view evidence generation as a longer-term endeavour over 10, 20 or even 30 years by sequencing a range of observational (exploratory) research studies until narrower combinations of two interventions become testable with more robust designs (whether through randomised trials, or other designs).
 - 4 Applying a suitable design to assess cost-effectiveness, drawing on cost–consequence analysis, cost-apportionment, cost–utility analysis or cost–benefit analysis (Masset *et al.*, this *IDS Bulletin*). However, because of the challenge of applying such techniques for purposes beyond their original intention – and often because of the lack of suitable comparators – more attention than usual needs to be paid to the interpretation of findings (see Acharya *et al.*, this *IDS Bulletin* for details).

By designing specific interventions around combinations that can be robustly evaluated in the four points described above – and combining this with observational research to provide explanatory power – this provides a way to increase our knowledge of the interaction of two or more activities leading to two or more outcomes. While it may not fully answer the immediate need for a mid-range theory of synergy (as emphasised by Masset, this *IDS Bulletin*), over time it will go a long way to incrementally building up such a theory. As FHI 360's *Catalyzing Integration* series shows, there is an emerging rich vein of evidence around what works and why in integration: from 'water, sanitation and hygiene with education interventions', 'agriculture with nutrition', 'governance, agriculture and food security', and 'climate change, agriculture and food security'.¹⁰

And finally, there is one aspect that such an approach to evaluating integration will overlook. As Jupp *et al.* (this *IDS Bulletin*) stress in their article, there can be significant disconnects between what an integrated project aspires to and how people actually experience and perceive change. Indeed, knowing in advance with enough certainty what to test empirically remains a challenge.¹¹ This may be none-the-more-so for integrated projects, as Masset (this *IDS Bulletin*) illustrates, with many aspiring to grand theory beyond the project modality.¹²

Plus, whereas IRD projects of the past attempted to move beyond agricultural productivity by adding a package of basic social services, the SDG era now poses an even greater challenge: not only to consider the interaction between different *sectors* (health, education, agriculture, and so on) but also to take it up another level and consider the system-wide effects (how the market economy can be more equitable and sustainable). Systems-based evaluation has some way to go, but climate change and market systems research, for example, are starting to address this challenge (e.g. van den Berg and Cando-Noordhuizen 2017) through modelling and prediction or risk assessment. Further research and testing are needed to consider how the aspects of uncertainty and complexity inherent in most integrated projects are better assessed and better understood in the future.

Notes

- * This issue of the *IDS Bulletin* was prepared as part of the impact evaluation of the Millennium Villages Project in northern Ghana, 2012–17, funded by the UK Department for International Development (DFID) (www.dfid.gov.uk). The evaluation was carried out by Itad (www.itad.com) in partnership with IDS (www.ids.ac.uk) and PDA-Ghana (www.pdaghana.com). The contents are the responsibility of the evaluation team and named authors, and do not necessarily reflect the views of DFID or the UK Government.
- 1 Honorary Associate at IDS and Director of Technical Excellence, Itad, Hove, UK.
 - 2 Research is ongoing in this area, such as the Sussex Sustainability Research Programme (www.sussex.ac.uk/ssrp/research) and the International Science Council (e.g. ICSU 2017). The research demonstrates that the SDGs are strongly interconnected and in ways that often have not yet been fully explored.
 - 3 For example, FHI 360 has been conducting research on integrated approaches through projects such as ‘Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES)’, as well as producing evidence maps and synthesising evidence of integrated solutions (www.fhi360.org/expertise/research-integrated-development). Similarly, Locus, a coalition of international development organisations, is focused on integrated approaches to development (<https://locus.ngo/resources>).
 - 4 In July 2006, the independent evaluation functions of the World Bank were integrated into a single unit, the Independent Evaluation Group.
 - 5 www.fhi360.org/resource/integrated-development-evidence-map.
 - 6 Similarly, Masset *et al.* (this *IDS Bulletin*) define it in their article as a joint production through multiple activities, with two activities affecting the same two outcomes.
 - 7 For example, the micro–macro paradox of achieving effectiveness at an intervention-level (solar panels) but no impact on transformative change at a systems-level (a low-carbon economy) (van den Berg and Cando-Noordhuizen 2017).

- 8 The idea of nesting different evaluation designs and methods (at multiple levels, potentially in a hierarchy) to address different parts of the evaluation (Stern *et al.* 2012, drawing from Lieberman 2005).
- 9 Indeed, one of the initial designs for the MVP evaluation in northern Ghana considered having different treatment arms (with different combinations of interventions to be tested) (Masset *et al.* 2013).
- 10 www.fhi360.org/resource/integrated-development-tools.
- 11 This is what Rogers distinguishes as truly complex, rather than just complicated: 'Complicated interventions that have many components pose challenges to evaluations, given the limited number of variables that can be identified and empirically investigated. But it is complex interventions that present the greatest challenge for evaluation and for the utilization of evaluation because the path to success is so variable and it cannot be articulated in advance' (2008: 31).
- 12 For the MVP, this was by using project modality to achieve sufficient synergy through doing everything together that it would enable the poor (within and beyond the project area) to break out of a cycle of poverty on a longer-term basis.

References

- Ahner-McHaffie, T.W.; Guest, G.; Petruney, T.; Eterno, A. and Dooley, B. (2018) 'Evaluating the Impact of Integrated Development: Are We Asking the Right Questions? A Systematic Review', *Gates Open Research* 1.6: 1–23, <http://dx.doi.org/10.12688/gatesopenres.12755.2> (accessed 13 July 2018)
- Barnett, C. and Eager, R. (2017) 'New Frontiers for Evaluation in a Fast-Changing World', in R.D. van den Berg, I. Naidoo and S. Tamondong (eds), *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*, Exeter: International Development Evaluation Association (IDEAS)
- Barnett, C. *et al.* (2018) *Impact Evaluation of the SADA Millennium Villages Project in Northern Ghana: Endline Summary Report*, commissioned by the UK Department for International Development, Hove: Itad
- Befani, B.; Ramalingam, B. and Stern, E. (2015) 'Introduction – Towards Systemic Approaches to Evaluation and Impact', *IDS Bulletin* 46.1: 1–6, <http://bulletin.ids.ac.uk/idsbo/article/view/119> (accessed 3 October 2018)
- van den Berg, R.D. and Cando-Noordhuizen, L. (2017) 'Action on Climate Change: What Does It Mean and Where Does It Lead To?', in J.I. Uitto, J. Puri and R.D. van den Berg (eds), *Evaluating Climate Change Action for Sustainable Development*, Cham: Springer International Publishing
- Clemens, M.A. and Demombynes, G. (2013) *The New Transparency in Development Economics: Lessons from the Millennium Villages Controversy*, CGD Working Paper 342, Washington DC: Center for Global Development
- Clemens, M.A. and Demombynes, G. (2011) 'When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages', *Journal of Development Effectiveness* 3.3: 305–39

- ICSU (2017) *A Guide to SDG Interactions: From Science to Implementation*, Paris: International Council for Science
- Jimenez, E. and Puri, J. (2017) 'The Wicked Cases of Education and Climate Change – The Promise and Challenge of Theory-Based Impact Evaluations', in R.D. van den Berg, I. Naidoo and S. Tamondong (eds), *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*, Exeter: International Development Evaluation Association (IDEAS)
- Le Blanc, D. (2015) *Towards Integration at Last? The Sustainable Development Goals as a Network of Targets*, DESA Working Paper 141, New York NY: United Nations Department of Economic and Social Affairs
- Lieberman, E. (2005) 'Nested Analysis as a Mixed-Method Strategy for Comparative Research', *American Political Science Review* 99.3: 435–52
- Masset, E.; Acharya, A.; Barnett, C. and Dogbe, T. (2013) 'An Impact Evaluation Design for the Millennium Villages Project in Northern Ghana', *Journal of Development Effectiveness* 5.2: 137–57
- Mitchell, S. *et al.* (2018) 'The Millennium Villages Project: A Retrospective, Observational, Endline Evaluation', *The Lancet* 6: e500–13, www.thelancet.com/action/showPdf?pii=S2214-109X%2818%2930065-2 (accessed 2 October 2018)
- Picciotto, R. (2015) *The 5th Wave: Social Impact Evaluation*, New York NY: Rockefeller Foundation Evaluation Office
- Rogers, P.J. (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', *Evaluation* 14.1: 29–48
- Sachs, J.D. (2018) 'Lessons from the Millennium Villages Project: A Personal Perspective', *The Lancet* 6.5: PE472–E474
- Sachs, J.D. and McArthur, J.W. (2005) 'The Millennium Project: A Plan for Meeting the Millennium Development Goals', *The Lancet* 365.9456: 347–53
- Schwandt, T.; Ofir, Z.; Lucks, D.; El-Saddick, K. and D'Errico, S. (2016) 'Evaluation: A Crucial Ingredient for SDG Success', *IIED Briefing*, London: International Institute for Environment and Development
- Stern, E.; Stame, N.; Mayne, J.; Forss, K.; Davies, R. and Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, London: Department for International Development (DFID)
- Thomas, V. (2017) 'Evaluating Sustainable Development', in R.D. van den Berg, I. Naidoo and S. Tamondong (eds), *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*, Exeter: International Development Evaluation Association (IDEAS)
- UN Millennium Project (2005) *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals*, Report to the UN Secretary-General, Millennium Project, London: Earthscan Publications
- Wanjala, B.M. and Muradian, R. (2013) 'Can Big Push Interventions Take Small-Scale Farmers out of Poverty? Insights from the Sauri Millennium Village in Kenya', *World Development* 45: 147–60
- Wiesen, C. and Prokop, M. (2015) 'From Millennium Development Goals to Sustainable Development Goals: Challenges and Opportunities for the Evaluation Community', paper presented at the Fourth International Conference on National Evaluation Capacities, Bangkok, 26–30 October

This page is intentionally left blank