THE MILLENNIUM VILLAGES: LESSONS ON EVALUATING INTEGRATED RURAL DEVELOPMENT

IDS Bulletin
Transforming Development Knowledge

Institute of Development Studies

Volume 49 | Number 4 | September 2018

IDS Bulletin The *IDS Bulletin* is an open access, peer-review journal exploring emerging international development challenges. It is published bi-monthly and is the flagship publication of the Institute of Development Studies, a leading global institution for research, teaching and learning, and impact and communications, based at the University of Sussex. Progressive economic, social and political change for everyone needs new kinds of action and relationships, shaped by new kinds of research and engagement. The *IDS Bulletin* aims to transform development knowledge, through its unique thematic issues developed by global learning partnerships that bridge academic, practice and policy discourse.

Publishing Manager/Bulletin Editorial Coordinator Alison Norwood

Production Editor Beth Richard

Marketing and Production Officer Gary Edwards

Publisher The *IDS Bulletin* is published by the Institute of Development Studies, Library Road, Brighton BN1 9RE, UK.
Tel: +44 (0)1273 606261.

Cover photo Women and girls collecting water from a well in the village of Nyologu, Northern Region, Ghana.

Photographer Nyani Quarmyne/Panos.

Disclaimer The Publisher and Issue Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher and Issue Editors, neither does the publication of advertisements constitute any endorsement by the Publisher and Issue Editors of the products advertised.

Copyright *IDS Bulletin* © Institute of Development Studies

This issue is published under a CC BY-NC-ND licence.

This is an Open Access issue 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

Online open access This journal is available online at bulletin.ids.ac.uk. Visit the site to search for articles and authors and register for table of contents email alerts.

Information for subscribers The *IDS Bulletin* is published in six issues per year.

Advertising enquiries Gary Edwards, Marketing and Production Officer, IDS Communications and Engagement Unit, idsbulletin@ids.ac.uk

IDS Bulletin © Institute of Development Studies 2018

www.ids.ac.uk
IDS is a charitable company limited by guarantee and registered in England (No. 877338).

IDS Bulletin
Transforming Development Knowledge

IDS Bulletin
Transforming Development Knowledge
Volume 49 | Number 4 | September 2018

The Millennium Villages: Lessons on Evaluating Integrated Rural Development

Editor **Chris Barnett**

Foreword
Richard Longhurst vii

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

Acknowledgements

We acknowledge the very helpful comments and suggestions of the individuals who peer reviewed the articles in this *IDS Bulletin*. Plus, we extend our appreciation to Annette Brown, Holly Burke, and Emily Namey of FHI 360 (www.fhi360.org) and their work on integrated development.

The evaluation of complex development interventions like the Millennium Villages Project asks for creative, and often innovative, thinking about the mix of methods that can capture change, assess contribution, and make sense of contradicting evidence and heterogeneous impacts. We acknowledge the role of the Centre for Development Impact (CDI) in providing a space and community of practice to do so. CDI is a joint initiative between Institute of Development Studies (IDS), Itad, and the University of East Anglia. For information see: www.cdimpact.org.

Finally, we acknowledge the valuable assistance of Rachel Eager, Efi Alamanos, Alison Norwood, Beth Richard, and Janice Fitz-Graham.

Funder acknowledgements

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.

Vol. 49 No. 4 September 2018:
'The Millennium Villages: Lessons on Evaluating Integrated Rural Development'
DOI: 10.19088/1968-2018.155

Notes on Contributors

Arnab Acharya is an economist with a PhD in Microeconomic Theory and Development Economics from the University of Illinois. He also has a master's degree in Public Health from Harvard University. His publications range from theoretical economics to public health policy, development economics, and empirical political science. He has held academic posts in the US, the UK, and India. He was a Fellow at IDS and a senior lecturer at the London School of Hygiene & Tropical Medicine. He has worked on projects funded by the Department for International Development (DFID), the World Bank, and the United States Agency for International Development (USAID).

Chris Barnett (DPhil, University of Sussex) is the Director of Technical Excellence at Itad, and an Honorary Associate at IDS. He was formerly the Director of the Centre for Development Impact (CDI); a joint initiative exploring innovative ways to evaluate impact. Chris was the Project Director for the impact evaluation of the Millennium Villages Project in northern Ghana. His current research interests include non-experimental methodologies to evaluate impact and causality, power imbalances, and the ethics of evaluation, plus assessing the development impact of investments and other forms of innovative finance (prizes, blended finance, etc.).

Annette N. Brown, PhD is an economist with more than 20 years' experience combining both research and implementation in international development. She currently directs FHI 360's Research and Evaluation Strategic Initiative and is the Editor-in-Chief for the *R&E Search for Evidence* blog. Prior to joining FHI 360, Annette headed the Washington DC office of the International Initiative for Impact Evaluation (3ie) and served on its executive team. Earlier in her career, Annette worked at both for-profit and not-for-profit development implementers. She is the author or co-author of numerous articles published in public health and social science journals.

Holly M. Burke, PhD, MPH is a behavioural scientist in the Reproductive, Maternal, Newborn and Child Health (RMNCH) Division at FHI 360. Holly joined FHI 360 in 2002 where she designs and leads international health research studies. Her expertise is in (1) designing studies that incorporate both quantitative and qualitative research methods to expand understanding of contraceptive and HIV prevention behaviours, and (2) evaluating the impact of programmes, including health communication campaigns, and integrated economic strengthening and health interventions.

Mario Chen, PhD is a biostatistician with more than 20 years' experience working on a broad range of research projects including clinical trials, impact evaluations, systematic reviews, surveys, and

other observational studies. Mario is currently Associate Director of Biostatistics at FHI 360, where he oversees the biostatistical work for FHI 360's behavioural, social sciences, health services, and other programmatic research agenda. He is also the principal investigator for a National Institutes of Health (NIH) contract set up to support infectious diseases research. Mario is author or co-author of over 70 peer-reviewed publications in the fields of statistics, public health, and infectious diseases.

Tony Dogbe is a co-founder of Participatory Development Associates Ltd. He is best known in Ghana and internationally for the pioneering role he has played in participatory poverty assessments, including access of the poor to services (1995); consultations with the poor (1999); and participatory poverty and vulnerability (2009–10). He is an advocate for community-driven development, citizen-led advocacy, and the inclusion of participatory methodologies for impact studies. He was a member of the team that undertook the impact evaluation of the northern Ghana Millennium Villages Project (MVP). He managed the field teams and the processes for the qualitative studies.

Eva-Maria Egger is an applied economist in the Research and Impact Assessment division of the International Fund for Agricultural Development (IFAD). She researches on migration, climate change, and rural development. Before joining IFAD, Eva-Maria worked with the Migrating out of Poverty Research Consortium at the University of Sussex. She holds a PhD and MSc in Economics from the University of Sussex, UK, and a BA in Political Science and Economics from the Albert Ludwigs University of Freiburg, Germany.

Tom Hilton is a freelance development economist and monitoring and evaluation expert with some ten years' experience working in private sector development. Having previously worked with Itad and the Gatsby Foundation, he is currently based in Latin America, working on a range of projects relating to market systems development, impact investment, forestry, and climate change.

Dee Jupp, PhD has over 30 years of participatory development and qualitative research experience. She is passionate about people-centred research, and developing and fusing methods to open up spaces for people's voices, perspectives, and lived reality to be heard and acknowledged by policymakers. Since 2007, she has been particularly involved in using the Reality Check Approach (RCA), an immersive qualitative research method, and she currently advises the global group of RCA practitioners. She also promotes the use of people's own indicators and categories for assessing change.

David Korboe is an independent development facilitator with a passion for transforming the lives of vulnerable and marginalised groups. He has evaluated and supported development projects for over 25 years. He has also taught participatory methods and social change approaches

to development practitioners and led one of Africa's first participatory poverty assessments in the early 1990s.

Lisa C. Laumann leads FHI 360's United States Agency for International Development (USAID)-funded Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) Family Care project, focused on developing programme guidance on household economic strengthening to keep children and families together. Lisa's prior work has included programme and country leadership roles in Pakistan, Afghanistan, Indonesia, Haiti, Lebanon, and at the headquarters/global level with Save the Children, as well as time with the International Rescue Committee and the US Peace Corps. She holds a master's degree in Education and International Development (Stanford University) and a PhD in Comparative Education (University of California, Los Angeles).

Giulia Mascagni is a Research Fellow at IDS and Research Director at the International Centre for Tax and Development (ICTD), Brighton. Her research interests focus on tax administration and tax policy in low-income countries, particularly in African countries. She is also Research Associate at the Institute of Fiscal Studies, London.

Edoardo Masset is Deputy Director of the Centre of Excellence in Development, Impact and Learning (CEDIL), London. Before joining CEDIL, Edoardo was Deputy Director and head of the London office of the International Initiative for Impact Evaluation (3ie), which he joined after working for seven years as a Research Fellow at IDS. Edoardo is an agricultural and development economist with extensive experience conducting impact evaluations of development interventions. His core research interests include rural development, poverty and inequality, child nutrition, and the analysis of household surveys.

Emily Namey is Associate Director of FHI 360's Behavioral, Epidemiological and Clinical Sciences division within the Global Health, Population and Nutrition unit. She has over 15 years' experience designing, implementing, and disseminating qualitative and mixed methods research across a range of topic areas including infectious disease, maternal and reproductive health, bioethics, economic strengthening, and child protection. She has conducted qualitative and mixed methods research/evaluation training courses in more than a dozen countries and has co-authored several methodological textbooks. Emily has an MA in Applied Anthropology, with an emphasis on public health and medical anthropology, from Northern Arizona University.

Amrita Saha is an IDS post-doctoral researcher and London School of Economics (LSE) Visiting Fellow with over eight years' experience working on the political economy of development. Her research explores the empirical political economy of trade policy, agricultural commercialisation, innovation, and inclusive structural change.

Amrita holds a PhD in Economics (University of Sussex). Her past research roles have been with the Commonwealth Secretariat, the Ministry of Commerce, India, the World Trade Organization (WTO), and the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). She is particularly interested in working on political economy issues for India and developing countries.

Foreword

Over the last 20 years, evaluation methods have broadened in scope, approaches, and hence value for decision-making for policymakers. This *IDS Bulletin* takes a big step forward to show how integrated multisectoral development – not a new area, but one which has all too easily been glossed over – can be given a new lease of life in today's contemporary debates. As the *IDS Bulletin* points out, poor people are unlikely to consider life as a series of disconnected sectoral issues or challenges, so there is a strong argument for more connected approaches.

In evaluating and investigating this challenge, this issue reflects on: the use of quantitative methods, a better understanding of synergy, the costs of integration, the added value of mixed methods observational research for integrated projects, and the challenge of dealing with multiple outcomes on different timescales, with different activities and inputs.

The significance of this *IDS Bulletin* is how the evidence, ideas, and approaches expressed here can illuminate one of the great mysteries of development research and practice. This is why so often the policy conclusions from research and evaluation emphasise the need for integration, coordination, addressing complexity, holistic approaches, exploitation of synergies (many peoples' favourite), and now the SDG-generating favourite, inter-connectedness, without much evidence as to how to get there. All of these outcomes are often stated as being intrinsically 'good things'. But are they and how do we know? In practice, do the efforts, transaction costs, and requirements of interdisciplinarity approaches to integrate outweigh the benefits? What is the best way to do this? A good first step would be to address the incentives that encourage silo-working.

This *IDS Bulletin* bravely takes on trying to find out. Case studies are presented on historical evidence, and detailing the pros and cons of diverse evaluation methodologies. Observations are drawn from case studies from the integrated rural development programmes of the 1970s, the Millennium Villages Project, and projects in Uganda and South Africa that implemented integrated family, youth, and economic strengthening.

Integrated development is messy to analyse and evaluate in order to be confident that it meets the needs and realities of poor people. This *IDS Bulletin* shows that integrated projects can be effective, but that the actual mechanisms compared to alternative approaches need more investigation.

This *IDS Bulletin* will suit those who enjoy being pulled up sharp by convincing evidence.

Richard Longhurst
Institute of Development Studies, University of Sussex

This page is intentionally left blank

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

Chris Barnett[1]

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 unwieldly 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

Integrated Development, Past and Present*

Edoardo Masset[1]

Abstract Recent years have witnessed a renewed interest in integrated rural development (IRD) projects, which were a common feature of international development in the 1960s and 1970s. In this article we critically review the literature on past IRD with the goal of informing current practice. We identify two key narratives in the IRD literature: (1) IRD projects were designed to exploit complementarities and synergies between development interventions, and (2) the administrative complexity of IRD projects prevented their successful implementation. We argue that the first narrative is not grounded in a solid theory of how IRD works, and that the second is largely based on a body of evidence which is wide but not rigorous. We show that some recent IRD experiences have been successful and conclude that future IRD evaluations need a novel conceptualisation of synergies and greater attention to the characteristics of implementation and cost-effectiveness.

Keywords: integrated development, Millennium Villages, integrated rural development, IRD, multi-sector, synergy, impact evaluation, complexity.

1 Background

Recent years have witnessed a surge of interest in multi-sectoral poverty reduction interventions. The project sites under the Millennium Village Project (MVP) discussed in this issue of the *IDS Bulletin* is one example. The MVP was conceived to show that the Millennium Development Goals (MDGs) could be achieved in rural Africa at a small geographic scale and at relatively small cost through interventions in multiple sectors (Sanchez *et al*. 2007). Reminiscent of theories of poverty traps popular in development economics (see, for example, Azariadis and Stachurski 2005), the hypothesis underlying the MVP was that simultaneous interventions in multiple sectors could raise living standards over a threshold level that would bring villages onto a sustainable development path, thereby breaking the poverty trap. The non-profit human development organisation FHI 360 has recently developed a research programme to identify the multi-sector and

© 2018 The Author. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.159
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'**; the Introduction is also recommended reading.

integrated programmes which are the most powerful and effective.[2] The Partnership for Economic Inclusion (PEI) hosted at the World Bank was recently established to accelerate and scale up the poverty graduation approach initiated by the non-governmental organisation (NGO) BRAC: a 'coordinated multi-sectoral big-push intervention' to address extreme poverty.[3] The poverty graduation approach consists of a package of cash transfers, training, asset transfers, and financial inclusion interventions, which was successfully tested through a multi-country evaluation by the Abdul Latif Jameel Poverty Action Lab (J-PAL) in nine countries (Banerjee *et al.* 2015).

These multi-sectoral interventions with multiple goals are well aligned with the recent transition from the MDGs to the Sustainable Development Goals (SDGs). It has been observed (Le Blanc 2015) that the SDGs are strongly interconnected so that impact in one sector may produce second round impacts in other sectors. These interlinkages between outcomes invite the international community to consider development as a system of trade-offs and interdependencies. The interconnections between goals may enable more integrated policies and easier considerations of synergies and trade-offs across SDG areas. All this points to what seems to be a renovated interest by the international research and policy community in multi-sectoral integrated poverty reduction interventions. These interventions, however, are not new. Very similar projects were implemented on a massive scale in developing countries by NGOs and international organisations starting in the 1960s. Integrated rural development (IRD) projects and small-area programmes dominated international aid assistance in the 1970s. These initiatives were abandoned in the early 1980s and the re-proposition of similar interventions nowadays has led some to argue that the international development community is unable to learn from past experience, and for this reason is bound to repeat the same mistakes again and again (White 2015).

In this article we critically review the literature on past and recent integrated development programmes. Our primary goal is to draw lessons from past experience that can inform the design, implementation, and evaluation of better integrated development programmes. We start with a summary of existing reviews of past integrated development programmes, followed by a discussion of synergies and of the available evidence of their effectiveness. We then review the success of recent integrated development programmes and we conclude with some reflections on the design and evaluation of future integrated interventions.

2 Old integrated rural development

There is a large literature on integrated development programmes. The World Bank, one of the major funders of integrated development programmes, has been particularly prolific. In this article, we mostly rely on existing reviews of this large literature, particularly on two reviews conducted by the Overseas Development Institute (ODI) (ODI 1979;

Buse, Ludi and Vigneri 2008), the review by Baah-Dwomoh (2016), and a very comprehensive review by the Operations Evaluation Department (OED) of the World Bank (World Bank 1988), and on other papers quoted by the same reviews.

IRD projects were very popular in the 1970s among major development agencies such as the World Bank, the African Development Bank, the Ford Foundation, the Rockefeller Foundation, the Swedish International Development Cooperation Agency (Sida), and the International Maize and Wheat Improvement Center (CIMMYT), among others. Two major conferences have been singled out as pivotal in a major shift towards IRD by major donors (Ruttan 1984). In 1971, a major symposium was held at the Food and Agriculture Organization of the United Nations in Rome on 'Agricultural Institutions for Integrated Rural Development' and in 1973, the World Bank president Robert McNamara made a speech to the Board of Governors in Nairobi in which he made a pledge to fight world poverty through an integrated approach to rural development. Since then, integrated rural development has become an increasingly important focus of multilateral and bilateral development agencies and NGOs.

Two main reasons are said to explain the popularity of IRD projects in the 1970s: the widespread persistence of poverty in rural areas and the perceived need to simultaneously address multiple constraints to economic growth. Development efforts in the 1950s and 1960s were mainly devoted to the promotion of industrialisation and community development following the theoretical thinking and the development discourse of the time. By the end of the 1960s, however, the persistence of poverty in rural areas suggested that these policies had largely failed and that entire segments of the rural population had been left behind by development interventions. It was felt that deprived areas were in need of special programmes to improve productivity and that a package of basic services had to be provided. At the same time, it was believed that poverty could not be addressed by simply promoting agricultural development. Poor people's opportunities, it was thought, were limited by multiple constraints in infrastructure, health, and education that needed to be addressed simultaneously. IRD projects promised to address these multiple constraints exploiting synergies and complementarities across sectors.

References to synergies and complementarities were common in the discussion of IRD projects of the time and they recalled theories of the 'big push' put forward by the 'high development theories' of the 1950s (Krugman 1995). Authors such as Rosenstein-Rodan (1943), Hirschman (1958), and Leibenstein (1957) had interpreted development not as a gradual process but the result of a radical transformation of the economy through simultaneous changes in all sectors. According to these authors, economic progress would require simultaneous growth in all sectors through the operation of forward and backward linkages and economies of scale. By the early 1970s, these development theories were

completely discredited in academic and policy circles, but the notion of synergies and complementarities across sectors lingered on in the practice of IRD projects, though the precise nature of the hypothesised interactions across sectors such as agriculture, health, and education was never investigated or described.

The practice of integrated rural development took many different forms including small-area interventions, packages of agricultural interventions, and truly holistic programmes. The pioneering village development project run by the Pakistani Academy for Rural Development at Comilla in today's Bangladesh was a great source of inspiration for the designers of IRD projects (Ruttan 1984). The Academy was established in 1959 as an experimental training station and it evolved into a holistic development programme based on the cooperation between local authorities and communities and coordinating activities in agriculture, water, health, and education. Observers of the time reported that the project had been successful in developing cooperatives and in increasing farmers' productivities. Following the Comilla example, similar integrated projects were established in other countries such as the project in Puebla (Mexico) and in Lilongwe (Malawi) (World Bank 1988).

IRD projects shared some common characteristics. They all had a focus on agriculture and the goal of increasing crop productivity and agricultural incomes. This was normally pursued through a package of interventions including cooperative development, credit access, input delivery, access to roads, and markets. But IRD projects went well beyond agricultural interventions and were often multi-sectoral, including interventions in health, education, and infrastructure. They relied on a centralised unit that coordinated and carried out activities in different areas. Since programmes were often implemented in deprived areas lacking the institutional structure and skills to provide this coordination, the coordination units were established anew and relied on highly skilled technical staff and expatriates.

IRD projects became the standard approach to rural development in the 1970s. For example, the World Bank approved 227 such projects between 1975 and 1989, though it never used the term IRD explicitly and preferred the term 'area development projects' (World Bank 1988). IRD projects fell out of favour in policy circles in the early 1980s with the emergence of the Washington Consensus and of a new ideology hostile to agricultural development and public sector interventions. By the late 1980s, all IRD projects had been abandoned, though elements of IRD survived in Community-Driven Development (CDD) and in projects promoted by some multilateral agencies such as the International Fund for Agricultural Development (IFAD) and by NGOs.

The great popularity of IRD in the international development community and its later sudden demise appears to be related more to changes in the political environment than to an appreciation of evidence

of their effectiveness. The widespread adoption of IRD in the 1970s reflected a major shift in development policy, from a concern with economic growth to the goal of reducing income disparities, fighting poverty, and meeting the basic needs of the poor (Rondinelli 1979). The World Bank in the 1970s made unprecedented financial disbursements in rural development in an effort to fulfil the goal of fighting rural poverty inaugurated by its President, Robert McNamara, often without proper planning and evaluation (World Bank 1988). Similarly, in the early 1980s, the new development policy of structural adjustment drastically reduced resources for public investments and poverty alleviation programmes (World Bank 1993), mostly on political and ideological grounds.

A vast literature on IRD projects was produced in the 1970s and narrative reviews of this literature have generally concluded that the evidence against the effectiveness of IRD projects was overwhelming. However, while skimming this literature, we cannot avoid a sense of missed opportunities. In little more than a decade, hundreds of large-scale projects were implemented in different forms, by different agencies in an incredibly wide variety of contexts, but without an evaluation system in place to learn from these experiences. While a consensus has emerged that these projects were not successful, a number of questions remain unanswered. For example, to what extent did projects fail because of a fundamental failure in design or because of implementation problems related to the absence of adequate political and administrative support? IRD projects were implemented following early successful experiences, but what were the contextual characteristics, of implementation, implementing agencies, and beneficiaries, that influenced their success? Could IRD interventions have been more effective had they adopted a more decentralised and participatory implementation approach? In this article, we do not develop these questions further, but instead focus on two key themes.

The literature on integrated rural development is dominated by two key narratives. The first narrative is that there are gains in integrating interventions across different sectors such as agriculture, health, and education. The reason is that integration exploits complementarities and synergies: impacts in different sectors, such as agriculture, education and health, reinforce each other. The second narrative is that integration, though theoretically appealing, does not really work. By the early 1980s, a consensus emerged among academics and policymakers that IRD projects were impractical and unfeasible. Observers had reached the conclusions that the complexity of integrating interventions outweighed the gains produced by synergies. We now discuss these two narratives more in detail.

3 Synergies and complementarities

All reviews and the documents of the time made reference to synergies and complementarities as a main motivation of IRD projects. However, though synergies and complementarities were always mentioned, their mechanisms of operation were never explicitly spelled out. They were

introduced as a sort of obvious and known fact of development policy without further elaboration. One exception is Brinkerhoff (1981) who discussed the theoretical underpinnings of integrated rural development in some detail. Development, he argued, was the result of complex and multifaceted interactions between interlinked economic, physical, political, and social phenomena. It should be natural, therefore, that development efforts should try to replicate this process through integrated interventions. Development should be unlocked by the emergence of complementarities and synergies across sectors.

In economics, complementarities refer to the fact that the optimal use of one input may require the use of another input. For example, fertiliser alone may not increase agricultural yields and needs to be complemented by water. Indeed, a combination of fertiliser, new seeds, and water made the 'green revolution package' that increased agricultural yields in Asia. We can extend the concept of complementarity in the production of one crop to complementarity in agriculture. For example, farmers may need a combination of a green revolution package, training, roads, and access to market to sustainably increase their incomes. Moving a step further, integration can occur across sectors as well as within sectors. We can think of a higher order integration across agriculture, infrastructure, health, and education. It could be argued, for example, that agricultural interventions need a minimum level of infrastructural development to be successful, or that behavioural change in health practices needs a population with a minimum basic level of education. This point was explicitly made by Adelman, Morris and Robinson (1976) who concluded a review of anti-poverty policy stating that single-policy interventions do not have lasting effects and that different interventions should be implemented simultaneously in different sectors to have a sizeable and long-lasting impact on rural poverty.

Synergies consist of mutually reinforcing outcomes, independently of whether activities are complementary in the sense defined above. For example, empowered women may start a business activity more confidently, which results in higher incomes and further increase in empowerment. Synergies can also be the result of economies of scale, as they are commonly understood in economics, that emerge from increasing the scale of production. Such economies of scale can be generated by integrating the provision of services at a larger scale. For example, joint planning and service use by a single integrated entity operating in different sectors may save resources and avoid duplications. In other words, synergies and complementarities of IRD projects can be thought of as a technology whose joint output is larger than outputs separately produced:

$$y(x_1,\ldots,x_n) > y(x_1) + y(x_2) + \ldots + y(x_n)$$

Output generated by the joint production of outcomes (on the left-hand side) in *n* different sectors is larger than the output produced separately and independently in each sector (on the right-hand side).

Synergies and complementarities are terms borrowed from economics and their frequent use may suggest that IRD was informed by economic theory. The similarity between the language of economics and the language of IRD, however, is only apparent. The economic theory of the firm is formulated for single production units using several inputs to produce a single output, not many. Complementarities in economics refer to combinations of inputs, not outputs. Economies of scale in economics emerge at the expansion of the input scale, not through output interactions. Even the economics concept of economies of scope refers to economies of scale in the production process generated by the simultaneous production of a variety of outputs. It does not refer to outputs affecting each other. To complicate matters, unlike output considered by economic theory, development outcomes such as education and health cannot be easily priced and monetised in a single unit. This complicates the task of optimising inputs for output maximisation discussed by economic theory, as overall output of an IRD project is not easily defined.

Finally, not all interactions between outputs are mutually reinforcing. There are also conflicting outcomes that need to be taken into account. For example, promoting agricultural production may lead to higher labour demand and less schooling, thus acting against education goals rather than in their favour. The literature on IRD contains limited discussion of these issues, and the economics literature is not of much help either. The notion of synergies and complementarities seems to rely on grand development theories popular at the time which rely on the concept of economies of scale and that favoured big-push interventions which were similar to those proposed by IRD projects. But 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.

4 IRD did not work

The second narrative dominating the IRD literature is that IRD projects did not work. It is clear that this type of project was abandoned by the late 1980s and treated with scorn and contempt by most development agencies. The reasons for this are many. First, IRD projects did not deliver the expected results, which were, admittedly, very ambitious. Yields and rural incomes did not increase significantly and poverty persisted. This was found by most evaluations of IRD interventions, a point to which we will return later. Some difficulties originated at the design stage (Baah-Dwomoh 2016). IRD projects were often designed using a top-down approach. Apparently, projects were designed with scarce knowledge of the context and the solutions proposed were often inadequate. This was more common for the agricultural packages such as new crops or cropping systems that were not always feasible or even desirable in the particular context in which they were promoted. Projects often included a simple package of inputs (fertiliser and improved seed) to increase agricultural productivity. No consideration was given to dryland areas, farmers' risk, adaptation

of technology to local context, or development of local and more appropriate technologies. This problem was exacerbated by the reliance on staff with a technical background and with a poor understanding of poverty and social issues.

We are also told that IRD projects were often implemented in deprived areas with little potential for economic growth, and in economic environments hostile to agricultural development. The economic policies of the 1960s and 1970s, consisting of overvalued exchange rates, price controls, and tariffs on goods, were directed to favour urban areas and industrialisation against agriculture. In such policy environments, it was extremely difficult to stimulate agricultural production. Agricultural prices controlled by the state were unfavourable to producers. Producers' prices were kept artificially low by government policies and sometimes they would fluctuate in the international markets, thus harming farmers. Commercialisation of output was neglected, and problems related to availability of labour and land were also ignored.

By far the most significant problem of IRD projects was their complexity. IRD projects turned out to be very difficult to administer and implement because of difficulties in coordinating activities across sectors and agencies. Projects were implemented in an integrated way across sectors rather than sequentially or in parallel. This organisational complexity often resulted in delays. In addition, local institutions in marginalised rural areas were weak and not sufficiently skilled to carry out the activities. To obviate this problem new structures were established to administer the projects that were often staffed with expatriate and skilled personnel. This staffing was impossible to sustain once project funding had ended, nor could such organisations be scaled up to cover a larger area. This came as no surprise to those who had argued on theoretical grounds against integration well before IRD projects started, and who had questioned whether the integration of services also required the integration of the service provider (Ruttan 1984). In one interesting example reported by an OED report (World Bank 1988), staff interviews led to the conclusion that an integrated rural development project in Bangladesh of the cost of US$177 million would have been more effective if implemented as four separate projects. We are now told that IRD projects would be less expensive if implemented separately:

$$c(x_1,\ldots,x_n) > c(x_1) + c(x_2) + \ldots + c(x_n)$$

The cost of administering an IRD project was higher (left-hand side) than the cost of running the components of the project separately (right-hand side). This observation is interesting because it runs against the synergy argument made earlier on. One implication of the synergies described by the equation in Section 3 is that costs of administering an IRD project should decrease as the scale of the intervention or the number of sectors increases. But this second equation leads us to

consider that, on the contrary, costs increase with integration, and in principle the two statements cannot be true.

As for the evidence in support of the poor impact of IRD, this is totally non-existent. The evaluation methods used at the time have been described as underdeveloped and intuition was often used instead of data (World Bank 1988). Data on outcomes were never collected from comparable control areas. In fact, data were rarely collected at all, even in project areas, and they were often of poor quality when collected. Finally, when data were collected they were rarely analysed. In our review of the literature, we could not find a single piece of evidence on the impact of the interventions. The evaluations of IRD projects conducted by the World Bank were probably the most rigorous of the time and consisted of *ex ante* estimations of internal rates of return (IRR). Project impacts in the form of IRR were estimated before the projects were implemented and, sometimes, were calculated again after the projects ended, by revising the original projections with whatever piece of data was available. We learn that the average rate of return of World Bank projects calculated in this way was 10.4 per cent, just above the minimum threshold of 10 per cent, and that this was highly disappointing. It is difficult to decide what to make of the evaluation literature on IRD projects. On the one hand, all authors concur that IRD projects did not achieve their ambitious goals and we are willing to believe them. On the other hand, no evidence is ever presented in support of these statements and we are left somewhat unconvinced.

5 Recent integrated development

Despite the desertion of large-scale IRD projects in the 1980s, the idea that addressing rural poverty requires a holistic approach to development has resisted. Projects run by NGOs and development agencies often include integrated elements and some projects have embraced integration altogether. Examples of the latter include the Upper Mandrare Basin Development project in Madagascar run by IFAD between 2001 and 2009, which aimed at reducing poverty through a package of capacity building, local initiatives, financial services, and roads; and the World Bank-funded Southwest Poverty Reduction Project (SWPRP) implemented in the counties of Guanxi, Guizhou, and Yunan of rural China, which consisted of income-generating activities, reforestation, promotion of off-farm employment, rural infrastructure, rehabilitation of schools and clinics, construction of roads, and piped water supply.

Perhaps the most notable example of recent integrated development is the MVP discussed in this issue of the *IDS Bulletin*. Since 2006, the MVP has been implemented in ten countries of sub-Saharan Africa by the Millennium Promise, the Earth Institute at Columbia University, and the United Nations Development Programme (UNDP) to show that the MDGs could be achieved in rural Africa at a small geographic scale and at low cost through interventions in multiple sectors. According to the project designers (Sanchez *et al*. 2007), the achievement of the

MDGs in rural Africa was prevented by multi-sectoral constraints including low agricultural productivity, a high burden of infectious disease, and poor infrastructure. The hypothesis underlying the MVP was that simultaneous interventions in multiple sectors can raise the capital stock over a threshold level that will put the local economy on a sustained path of economic growth, thereby breaking the poverty trap. To fulfil these goals, the MVP made concerted investments in agriculture, health, education, and infrastructure.

Another example of modern integrated development is the ultra-poor graduation programme run by BRAC. The programme combines traditional social support initiatives such as cash transfers, and long-term support such as life-skills training, asset transfers, enterprise development, and saving and planning for the future. 'By addressing the social, economic and health needs of families simultaneously, these programmes provide holistic support to participants, as they climb the ladder of economic self-reliance into a sustainable future' (BRAC 2018). It is also worth mentioning that since 2014, FHI 360 has been running a research programme on integrated development in the belief that 'integrated approaches to the design, delivery and evaluation of programs have the potential to make an enduring difference in people's lives'.[4] The programme includes an evidence map and a theory of change of integrated development, several case studies, and examples of catalysing integration between water, sanitation, and hygiene (WASH) and education; agriculture and nutrition; and governance agriculture and food security.

Finally, it should be noted that many development projects are designed as packages of multiple interventions including social protection and community-driven development projects. Even when these packages are implemented in a specific sector, such as employment, nutrition, or governance, they include many different interventions in the belief that they are all required to achieve the stated goal and that they reinforce each other. For example, the Chars Livelihood Programme in Bangladesh included, among others, the following activities: growth monitoring and promotion of nutrition interventions, the construction of homestead plinths, the provision of sanitary latrines, and access to clean drinking water; asset and cash transfers, training, social mobilisation, and women's empowerment (Nisbett *et al.* 2016). In another example, the Tuungane project in the Democratic Republic of the Congo (DRC) promoted the establishment of community-level committees and village-level committees, and funded interventions in education, transportation, water, sanitation, and agriculture (Humphreys, Sanchez de la Sierra and van der Windt 2012).

The reader may wonder why integrated development is still popular, given the overwhelming evidence against its effectiveness. Apart from the already mentioned inability to learn from the past, it must be noted that despite the strong resemblance between new and old integrated development, there are also some major differences. Both old and new integrated development models consider development

to be complex and that there are synergies and complementarities across sectors. Proponents of both believe that problems are better addressed simultaneously and preferably by a specialised coordination unit run by a project, though it was noted in Section 4 that specialised units, established to supplement lack of local administrative capacity, introduce an element of complexity in implementation that prevents their short-term success and their long-term sustainability.

But there are also some major differences between old and new IRD. Recent integrated projects are much less ambitious. The IRD projects of the 1970s covered large geographic areas within a country with the goal of permanently eradicating poverty. As a matter of comparison, the World Bank-sponsored Upper Region Development Project ran for nine years between 1977 and 1985 covering three undeveloped regions of northern Ghana comprising 125,000 households and investing today's equivalent of US$240 million (World Bank 1987), while the northern Ghana MVP covered just two districts of northern Ghana, lasted four years, reached about 3,900 households, and invested a total of £11.5 million (US$16 million equivalent) in the area. Recent integrated projects also rely on implementing units that are more flexible.

The IRD projects of the 1970s established big implementation units that turned out to be inefficient and unsustainable, while modern integrated projects build simpler structures. Associated with a more agile management structure, there is a different approach to decision-making which is more inclusive and participatory in comparison to the classical top-down approach of early IRD projects. There are also differences in the type of interventions implemented. It is fair to say that much more is known today about agriculture, health, and education technology than was known in the 1970s, including in the science of managing development. Sanchez *et al.* (2007), for example, pointed out that, unlike old IRD projects, the MVP was designed using science and evidence-based technologies and practices that had been proved to work. Recent integrated projects are likely to be much better designed and to promote more effective policies than was the case in the 1970s.

Perhaps the biggest difference between new and old integrated development, however, is in the monitoring and evaluation of the interventions. Old IRD interventions were rarely, if ever, evaluated, while more recent experiences of integrated development have been often independently evaluated, sometimes in a very rigorous way. As a result, there is today a sizeable body of evidence on the effectiveness of integrated interventions, which allows a more informed and balanced judgement about their significance. An evaluation showed that the IFAD-sponsored project mentioned above made a significant impact on income and food security, through increased agricultural production (Baah-Dwomoh 2016). Conversely, the SWPRP in rural China improved short-term incomes but had no long-lasting effects on consumption and poverty (Chen, Mu and Ravallion 2009).

Two rigorous evaluations of the MVP found mixed results. Mitchell *et al.* (2018) collected data from MVP sites from ten different countries and found impacts on 30 of 40 'MDG-related' outcomes, particularly in the agriculture and health sectors. They conclude that, consistent with an integrated rural approach, the intervention had a favourable impact on all MDG areas. Our evaluation of the northern Ghana Savannah Accelerated Development Authority (SADA) MVP found less positive results, with only seven out of 29 MDG targets being reached and no discernible impact on two key indicators: poverty and child mortality (see Jupp and Barnett, this *IDS Bulletin*). A multi-country evaluation of the BRAC poverty graduation programme found long-lasting effects on poverty reduction (Banerjee *et al.* 2015), a result that was confirmed by further evaluations in other countries.

The evidence produced by evaluations of recent integrated development programmes suggests that they can be effective and have long-lasting poverty reduction effects. However, these evaluations are largely silent on the two main questions regarding integrated development discussed in this article: synergies and the cost of integration. It is not clear from these evaluations whether the projects were able to unlock the expected synergistic effects, nor is there a discussion of the added value of integrating the management of interventions across sectors. We will further develop these two points in the following concluding section.

6 Conclusions

Our experience in evaluating the northern Ghana MVP project and our reading of the literature on past and present integrated projects taught us that two elements are key for our understanding and for the successful design of future interventions: synergies and the cost of integration. Synergies are the fundamental motivation for integrating interventions across sectors, but the mechanisms behind the operation of synergies are not well understood. Projects often take the concept of synergies implicitly as obviously linked to multi-sector interventions. Alternatively, they make reference to some grand development theory, relying on economies of scale or complementarities such as, for example, theories of poverty trap or of the 'big push'. Impact evaluations of integrated interventions put a lot of effort in outlining the programme theory of the intervention to describe in detail how the intervention in each sector affects the target outcomes and what the interactions are between activities.

What is missing is a middle-level theory that lies between the all-explaining grand theories of development underpinning integration and the detailed programme theories of specific interventions (Davey *et al.* 2018). What is needed is an effort at theorising and conceptualising synergies occurring at the implementation and output level at a sufficient level of abstraction to be applicable to several contexts. It seems that synergies should ultimately emerge as the result of interactions between individuals and that theorisations should start by formulating behavioural

hypotheses at the origin of mutually reinforcing outcomes. Within the economics discipline, a possible starting point for such conceptualisations is the study of multi-input and multi-output production functions and joint production in agricultural economics (Chambers 1988) and the study of emergent phenomena in social interactions through the lens of complexity theory (Durlauf 2005, 2012).

A second major gap in our understanding of integrated development relates to the cost of integration. Few evaluations have conducted a cost effectiveness analysis of integrated interventions. We stressed how little is known about the cost of integrating interventions across sectors. The opinion of observers of old IRD projects is that this cost is so high as to outweigh any benefit obtained through synergies. Early critics of IRD projects (Ruttan 1984) did not argue against the possibility of generating synergies but doubted the advantages of running interventions in different sectors simultaneously. The question is whether synergies can be simply achieved through the coordination of interventions running in parallel in different sectors and being implemented by separate specialised agencies, rather than by costly implementation units striving to implement all activities simultaneously in all sectors.

Answers to questions about synergies and about the cost of integration can only be provided by evaluations that are at the same time rigorous and concerned with more than just effectiveness. We have suggested how synergies need to be better understood and conceptualised through the development of a mid-level theory of integrated interventions. The theory can be developed to a sufficient level of generality to be applicable to a variety of interventions and contexts. This theory will then generate a number of testable hypotheses that can be addressed by specific evaluation methods. These methods can include factorial designs, which include as many intervention arms as there are interventions and interactions of interventions (Ahner-McHaffie *et al.* 2017). Factorial designs are able to provide the full range of synergistic effects but are very complex to run and extremely expensive.

An alternative approach, which is strongly linked to the formulation of mid-level theories, consists of conducting mechanism experiments (Ludwig, Kling and Mullainathan 2011), whereby the causal chain of the intervention is fully unpacked and the evaluation tests the most uncertain and unknown links to compose our understanding of interventions as in a jigsaw puzzle. For example, rather than running a full factorial design, it could be extremely informative to test the interaction of just two or three interventions by selecting less-disadvantaged areas where other achievements predicted by the synergistic theory are already obtained. However, more interesting and informative tests are likely to be generated by the development of a middle-level theory of integrated development interventions, which should be the first step of a research programme to explore synergies of integrated interventions.

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 Centre of Excellence for Development Impact and Learning (CEDIL) at the London School of Hygiene & Tropical Medicine (LSHTM).

2 www.fhi360.org/expertise/research-integrated-development.

3 www.microfinancegateway.org/sites/default/files/announcement/pei-brochure.pdf.

4 www.fhi360.org/resource/integrated-development-tools.

References

Adelman, I.; Morris, C.T. and Robinson, S. (1976) 'Policies for Equitable Growth', *World Development* 4.7: 561–82

Ahner-McHaffie, T.W.; Guest, G.; Petruney, T.; Eterno, A. and Dooley, B. (2017) 'Evaluating Integrated Development: Are We Asking the Right Questions? A Systematic Review', *Gates Open Research* 1.6: 1–18

Azariadis, C. and Stachurski, J. (2005) 'Poverty Traps', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth Vol. 1A*, Amsterdam: Elsevier

Baah-Dwomoh, J.K. (2016) *Integrated Rural Development in Africa*, background paper for the African Transformation Report 2016: Transforming Africa's Agriculture, African Center for Economic Transformation (ACET)

Banerjee, A. *et al.* (2015) 'A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries', *Science* 348.6236: 772–88

BRAC (2018) *BRAC's Ultra-Poor Graduation Programme: An End to Extreme Poverty in our Lifetime*, Dhaka: BRAC Centre

Brinkerhoff, D.W. (1981) *The Effectiveness of Integrated Rural Development: A Synthesis of Research and Experience*, Washington DC: United States Agency for International Development (USAID)

Buse, K.; Ludi, E. and Vigneri, M. (2008) *Beyond the Village: The Transition from Rural Investments to National Plans to Reach the MDGs: Sustaining and Scaling Up the Millennium Villages*, London: Overseas Development Institute (ODI)

Chambers, R.G. (1988) *Applied Production Analysis: A Dual Approach*, Cambridge: Cambridge University Press

Chen, S.; Mu, R. and Ravallion, M. (2009) 'Are There Lasting Impacts of Aid to Poor Areas?', *Journal of Public Economics* 93.3–4: 512–28

Davey, C. *et al.* (2018) *Designing Evaluations to Provide Evidence to Inform Action in New Settings*, CEDIL Inception Paper 2, London: Centre of Excellence for Development Impact and Learning (CEDIL)

Durlauf, S.N. (2012) 'Complexity, Economics, and Public Policy', *Politics, Philosophy and Economics* 11.1: 45–75

Durlauf, S.N. (2005) 'Complexity and Empirical Economics', *The Economic Journal* 115.504: 225–43

Hirschman, A.O. (1958) *The Strategy of Economic Development*, New Haven CT: Yale University Press

Humphreys, M.; Sanchez de la Sierra, R. and van der Windt, P. (2012) *Social and Economic Impacts of Tuungane: Final Report on the Effects of a Community Driven Reconstruction Program in Eastern Democratic Republic of Congo*, New York NY: Columbia University

Krugman, P. (1995) *Development, Geography, and Economic Theory*, Cambridge MA: MIT Press

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 (UN DESA)

Leibenstein, H. (1957) *Economic Backwardness and Economic Growth*, New York NY: John Wiley

Ludwig, J.; Kling, J. and Mullainathan, S. (2011) 'Mechanism Experiments and Policy Evaluations', *Journal of Economic Perspectives* 25.3: 17–38

Mitchell, S. *et al.* (2018) 'The Millennium Villages Project: A Retrospective, Observational, Endline Evaluation', *The Lancet Global Health* 6.5: e500–e513

Nisbett, N. *et al.* (2016) *MQSUN Report. Impact Evaluation of the DFID Programme to Accelerate Improved Nutrition for the Extreme Poor in Bangladesh: Final Report*, Maximising the Quality of Scaling up Nutrition (MQSUN), Brighton: IDS

ODI (1979) *Integrated Rural Development*, ODI Briefing Paper 4, London: Overseas Development Institute

Rondinelli, D.A. (1979) 'Administration of Integrated Rural Development Policy: The Politics of Agrarian Reform in Developing Countries', *World Politics* 31.3: 389–416

Rosenstein-Rodan, P.N. (1943) 'Problems of Industrialisation of Eastern and South-Eastern Europe', *The Economic Journal* 53.210–11: 202–11

Ruttan, V.W. (1984) 'Integrated Rural Development: A Historical Perspective', *World Development* 12.4: 393–401

Sanchez, P. *et al.* (2007) 'The African Millennium Villages', *Proceedings of the National Academy of Sciences of the United States of America* 104.43: 16775–80

White, H. (2015) 'Fieldwork is Easy with Eyes Closed, Misunderstanding All You See… and Other Reflections: 3ie Annual Howard White Inaugural Lecture', *Journal of Development Effectiveness* 7.3: 375–87

World Bank (1993) *Area Development Projects*, OED Lessons and Practices 3, Washington DC: Operations Evaluation Department (OED), World Bank

World Bank (1988) *Desarrollo Rural – Experiencia del Banco Mundial*, Washington DC: Operations Evaluation Department (OED), World Bank

World Bank (1987) *Ghana – Upper Region Agricultural Development Project: Project Completion Report*, Washington DC: Operations Evaluation Department (OED), World Bank

This page is intentionally left blank

The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies*

Edoardo Masset,[1] Giulia Mascagni,[2] Arnab Acharya,[3] Eva-Maria Egger[4] and Amrita Saha[5]

Abstract Most development interventions are complex, comprising several interacting activities affecting multiple outcomes. Impact evaluations of such interventions are widespread, but the literature offers little guidance on how to assess the cost-effectiveness of such integrated projects. We review the literature that conducts cost-effectiveness analyses of multiple interventions alongside impact evaluations in low- and middle-income countries. Only seven studies are identified in areas as diverse as de-worming, school support, conditional cash transfers, early childhood development, and social funds. We find that none of the applied approaches can be effectively employed in all instances, though each of them can be applied in some special cases. Furthermore, none of the studies reviewed addresses output synergies. Given the rising numbers of impact assessments in development practice and their importance for policy, research needs to develop sound methods to assess the cost-effectiveness of integrated interventions.

Keywords: systematic review, methodology, cost-effectiveness analysis, complexity, multisector integrated development, synergy, cost–consequence analysis, cost apportionment, cost–utility analysis, cost–benefit analysis.

1 Introduction

Thanks to efforts of organisations such as, among others, the International Initiative for Impact Evaluation (3ie), the Abdul Latif Jameel Poverty Action Lab (J-PAL), and the World Bank, a large number of impact evaluations have been conducted in low- and middle-income countries in recent years. The latest figures for these countries suggest that in the past 20 years, more than 4,000 impact evaluations were undertaken (Sabet and Brown 2018). Impact evaluations have estimated the impact of development interventions in areas such as agriculture, education, health, infrastructure, and governance. Most development interventions under evaluation are complex, however, being composed of

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.160
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'**; the Introduction is also recommended reading.

several interacting activities affecting multiple outcomes. This integration across activities complicates the policymaker's task of using evaluation results to identify effective policies from those that are ineffective.

The policymaker's task to identify effective policies is further complicated by alternative policies aiming at the same goal. For example, what is the best way to reduce rural poverty: building roads, providing fertiliser subsidies, or conditional cash transfers? The analysis of impact alone is not sufficient here to determine effective policies. It is of importance to identify the costs of achieving the given impacts across policies. These choices can be informed by cost-effectiveness studies that consider the impact of the interventions in relation to their costs in a comparative way. Indeed, some impact evaluations collect data on project costs data and calculate cost-effectiveness ratios, the cost for obtaining one unit of the benefit outcome.

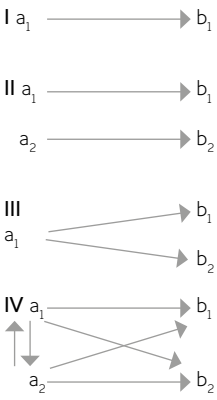
These cost-effectiveness analyses, however, face one major difficulty. Multiple activities in development projects also result in multiple outcomes, some of which are unintended. Multiple outcomes are not easily aggregated into a single index of effectiveness. It is not obvious that overarching welfare indices can be formulated and aggregated over different sectors. In relation to cost data, when there are multiple outcomes, we face the opposite problem: total project budgets cannot be easily disaggregated between the different project activities. As a result, it is often challenging to assign the cost of a project activity to its intended outcome. These are challenging methodological issues and reference books on cost-effectiveness analysis, such as Levin and McEwan (2001) and Drummond *et al*. (2005), offer little or no guidance as to how to assess the cost-effectiveness of complex interventions.

In this study, we review how researchers have conducted cost-effectiveness analyses of integrated development programmes within impact evaluations in low- and middle-income countries. The primary goal is to identify the best practices that are currently being used in order to assess the cost-effectiveness of complex interventions. A secondary goal is to map the existing cost-effectiveness literature of development programmes and identify gaps. Finally, we provide some recommendations for researchers conducting cost-effectiveness studies of complex interventions.

2 What are complex interventions?

We define complex interventions as interventions consisting of one or multiple activities and producing multiple outcomes. Conversely, we define a simple intervention as an intervention consisting of one activity and one outcome, and no unintended outcomes. Figure 1 illustrates the different cases considered. In Figure 1, a_1 and a_2 represent intervention activities while b_1 and b_2 represent the benefits/outcomes of these activities. We employ a maximum of two activities and benefits in order to simplify, but the exposition can be easily generalised to more than two activities and benefits.

Figure 1 A typology of complex interventions



Simple intervention: one activity and one benefit.
Example: insecticide-treated nets (a_1) and malaria incidence (b_1).

Disjoint interventions: activities are independent and there is a one-to-one match between activities and benefits.
Example: food supplementation (a_1), children's nutritional status (b_1), nutrition training (a_2), and nutrition awareness (b_2).

Intervention with multiple outcomes: one activity affects two outcomes.
Example: school feeding (a_1), school attendance (b_1), and nutritional status (b_2).

Complex package: two activities affecting the same outcomes and affecting each other.
Example: integrated rural development project promoting agricultural productivity (a_1) and education (a_2), with income (b_1) and enrolment (b_2) outcomes.

Source Authors' own.

Case I is a simple intervention. There is one activity (a_1) producing a single benefit (b_1). In the absence of unintended benefits, an incremental cost-effectiveness ratio can be calculated as the project budget (the cost of a_1) over the impact on the outcome (b_1), which represents the cost of obtaining a unit of the benefit for this particular intervention. For example, Bathia, Fox-Rushby and Mills (2004) compare the cost-effectiveness of two malaria control interventions: in-house residual spraying and insecticide-treated nets. The two interventions are randomly allocated at the community level and the prevalence of malaria before and after the intervention is assessed. Using project budget data, the authors calculate the average cost of averting one case of malaria for the two interventions. Insecticide-treated nets turn out to be more cost-effective by a large margin. Cost-effectiveness studies of simple interventions like this one do not present particular problems and will not be covered by our review. Case II is simply an extension of case I; it is presented to emphasise that many projects may have very separate distinct activities producing different outputs, and costs can easily be apportioned.

While cases I and II are simple, we believe they are also rare. Cases III and IV are complex and are much more common in international development. In case III, a single intervention achieves two outcomes; there is joint production through what can be thought of as a single activity. For example, school feeding programmes affect school attendance as well as the nutritional status of children. In this case, cost apportionment is not possible and cost-effectiveness ratios cannot be calculated to account for all project effects. In case IV, two activities affect the same two outcomes and also affect each other; there is joint production through multiple activities. For example, integrated rural development projects include, among others, activities promoting agricultural productivity and education. Both activities affect income and school attendance. The activities can also affect each other, for example, as more educated farmers are more likely to benefit from

Table 1 **PICOTs[6] inclusion/exclusion criteria for selecting studies**

	Included	Excluded
Population	Low-income, lower-middle-income, upper-middle-income	High-income, high-income OECD
Intervention	Complex programmes with multiple outcomes in different sectoral domains	Single-outcome programmes; programmes with multiple outcomes in the same domain; programmes with multiple outcomes evaluated only on one outcome
Comparison	Impact evaluations using experimental or quasi-experimental methods	Evaluations not employing experimental methods (RCTs) or quasi-experimental methods (regression discontinuity, matching methods, difference-in-differences)
Outcome	Multiple outcomes from different sectoral domains	Single outcomes; multiple outcomes within the same domain
Time	After 1999	Before 2000

Source Authors' own.

technical assistance. In this case, cost-effectiveness ratios can be calculated by cost apportionment but are difficult to interpret, as the cost-effectiveness ratio of each activity is not independent of costs incurred in the other activity.

3 Methodology: systematic review[7]

In this study, we conduct a systematic review of cost-effectiveness analyses of complex interventions in low- and middle-income countries. In our review, the term 'cost-effectiveness analysis' is equivalent to the term 'full economic evaluation' employed by Drummond *et al.* (2005: 9), a 'comparative analysis of alternative courses of action in terms of both their costs and consequences', which includes cost-effectiveness analysis, cost–utility analysis (CUA), and cost–benefit analysis (CBA).

3.1 Search

We searched published and unpublished literature from the following databases: Medline, ERIC, the Social Sciences Citation Index, Econlit, IDEAS/RePEc, the J-PAL website, the World Bank website, the International Food Policy Research Institute (IFPRI) website, and the 3ie repository of impact evaluations. We included in the search all studies produced in English that included either in the title or the abstract the following terms: cost analysis; cost-effectiveness; cost–utility; or cost–benefit.

3.2 Selection

We adopted the five inclusion and exclusion criteria listed in Table 1. First, studies of high-income countries were excluded. Second, we excluded simple programmes and only included complex multiple-outcomes interventions, as described in Figure 1. However, complex multiple outcomes interventions that were analysed only in relation to

one outcome were also excluded, as they do not offer any methodological insight. Third, we included only impact evaluations using experimental and quasi-experimental design: randomised control trials (RCTs), regression discontinuity designs (RDDs), matching methods, and difference-in-differences (DID) analyses. Fourth, we excluded studies looking at multiple outcomes that are different manifestations of the same latent construct, and we considered only multiple outcomes across different domains. For example, an intervention aiming at promoting women's empowerment and HIV prevention would be included, but an intervention aiming at improving school attendance and test scores would not be included. Lastly, we excluded all studies produced before 2000. This choice was made in the belief that few studies would be found before this date, given that the surge in impact evaluation studies in international development is a very recent phenomenon (Sabet and Brown 2018).

4 Findings

The systematic review resulted in only seven studies that fulfilled all criteria. This is surprising, as we find that an increasing number of impact evaluations (Sabet and Brown 2018) have not in turn produced comparable and credible CEA studies of complex projects. For such complex development projects, the limited results identify a critical knowledge gap where the policymaker's task of identifying effective policies remains constrained due to insufficient methods of identifying the costs of achieving given impacts across policies.

The detailed results of the search and selection process are summarised in Annexe 3. All the studies reviewed used effect sizes calculated using experiments (four randomised trials) or quasi-experiments (three propensity score-matching studies). The studies evaluated a wide range of development interventions: deworming (2); conditional cash transfers and food transfers (2); social funds (1); early childhood development (1); and preventative HIV school support (1). Costs were calculated using the 'ingredients approach' (all the components of the overall cost) in five cases and only two studies included 'social costs' (the cost of subsidising the deworming intervention and co-payments by the parents in the HIV preventative school support). Two studies also employed time discounting of benefits and costs (accounting whether a cost (or benefit) arises immediately or in the future).

4.1 Cost-effectiveness methodologies of the selected studies

To address the complexity of the interventions analysed, the studies took different approaches to assess the cost-effectiveness of the project under scrutiny. We review these approaches from the lens of our research question, i.e. whether and how the complexity challenge can be addressed with the applied method. To this end, we classified the cost-effectiveness approaches employed by the selected studies into four categories: cost–consequences analysis; cost apportionment; cost–utility analysis; and cost–benefit analysis. In what follows, we provide a brief description of each approach followed by a description and a critical appraisal of the studies employing this particular approach.

4.1.1 Cost–consequence analysis

In this approach, all project costs (c_1, c_2, ..., c_n) and benefits (b_1, b_2, ..., b_n) relating to project activities (a_1, a_2, ..., a_n) are presented in a table for each of two or more alternative interventions (Drummond *et al.* 2005). This approach has a number of advantages (Mauskopf *et al.* 1998). First, it is extremely simple, and costs and benefits presented in this way are easily understood by policymakers. Second, it is transparent as no implicit trade-offs are imposed on decision makers. Third, it allows a more flexible decision-making process. Decision makers at different levels of operations have different goals and the results presented by cost–consequence analysis allow them to apply their own preferences in any specific context in which decisions are made. The main disadvantage of cost–consequence analysis is that it is more effective when there are few interventions and few outcomes. In these cases, the most cost-effective intervention can be identified by finding the 'dominant' project, which outperforms the other projects on all outcomes. For example, an education intervention may be unambiguously superior to another one by being more cost-effective in terms of both mathematical and reading achievements (Levin and McEwan 2001). However, such comparisons made by isolating dominant projects can only be made when comparing similar interventions across other complex projects.

The dominant choice is more difficult when an intervention is superior to another one on one domain but inferior on another domain; for example, if an education project is more cost-effective at improving mathematics test scores but less cost-effective at improving reading skills. In this case, as well as in the case when there are many projects and many outcomes, it is likely that decisions based on cost–consequence analysis are biased by visual inspection, vote counting, or other salient data. This approach limits an understanding of the potential connections between costs and respective benefits that may be different across interventions. By presenting the intervention costs and benefits as separate, it oversteps the problem of linking them. We are therefore somewhat sceptical that this method sufficiently accounts for the complexity of integrated projects.

Our review found three applications of this approach (Ahmed *et al.* 2009; Hidrobo *et al.* 2012; Miguel and Kremer 2004). Hidrobo *et al.* (2012) compare the cost-effectiveness of three different implementation modalities of a conditional cash transfer programme in Ecuador. Impact estimates are obtained through a randomised trial for the following outcomes: consumption, calories, household diet diversity score, dietary diversity score, and food consumption score. Costs of achieving a 15 per cent increase in each outcome are obtained for each intervention modality and compared. In the calculation of the cost-effectiveness ratio, the full cost of each intervention modality is divided by each outcome after removing project costs that are common across the three modalities. The results are presented in a table which shows that the food modality is dominated by the other two. The voucher

modality is superior to the cash modality for all but one outcome. This study is a good illustration of the cost–consequences approach but presents a number of limitations. First, the project had other two objectives in addition to increasing food security, namely empowering women and reducing tension between refugees and the host population. However, the cost-effectiveness analysis is entirely focused on the food security outcomes. Second, the outcomes considered are highly correlated and could be perceived by policymakers as attributes of the same construct so that the information provided may be excessive. Third, the authors did not report confidence intervals and the differences observed between cost-effectiveness ratios of difference interventions may lie between those intervals.

Ahmed *et al.* (2009) compare the cost-effectiveness of four different social transfers interventions to poor women in Bangladesh. Project effects are estimated using propensity score matching (PSM), and cost-effectiveness ratios of the four interventions are compared in two domains: poverty (cost of increasing per capita daily calories intake by 100Kcal; cost of increasing household monthly income by 100 Taka; and cost of reducing extreme poverty by 1 per cent) and empowerment (cost of increasing women's participation in food decision-making; and the cost of increasing the percentage of women taking loans from non-governmental organisations (NGOs) by 1 per cent). Total project costs are used for all the outcomes. Two interventions clearly dominate the others in the income domain, while one intervention dominates all others in the empowerment domain. The dominant projects, however, are not the same across domains. The authors are careful not to draw conclusions from the analysis. This analysis suffers from the same problems highlighted in the previous study: cost-effectiveness is not measured for several other impacts of the intervention, the outcomes considered are highly correlated within each domain, and the point estimates do not include confidence intervals.

The study by Miguel and Kremer (2004) reports the use of four different cost-effectiveness approaches: a *health cost-effectiveness* approach, an *educational cost-effectiveness* approach, a *human capital investment* approach and an *externality* approach. Some of these approaches are applications of the cost–benefit analysis approach and will be discussed again below. However, the first three approaches together are an application of a cost–consequences approach. The authors estimate the impact of a randomised deworming intervention in rural areas of Kenya on school attendance, test scores (English, Mathematics, and Science), and parasitic worm infection. They calculate the cost per disability-adjusted life years (DALY) averted, the cost per additional year of schooling, and the increase in returns of education given the cost of treating a child. The cost per DALY is compared to a benchmark value for developing countries and the programme is found to be highly cost-effective. The cost per year of education is compared to cost-effectiveness ratios calculated from other interventions promoting primary education in Kenya, and it is found to dominate all alternative interventions.

Finally, the authors estimate a large increase in the net present value of wages. The combination of these results leads the authors to suggest that deworming is a highly cost-effective intervention in at least three different domains. This analysis has some limitations. First, the cost per treated child is obtained not from project data but from a similar project in Tanzania, which appears a bit arbitrary and *ad hoc*. Second, non-education interventions used as comparators are mostly hypothetical rather than real projects. Third, no confidence intervals for the cost-effectiveness ratios are reported. Finally, cost-effectiveness comparisons are not extended to test score outcomes for which the evaluation did not find a positive impact.

4.1.2 Cost apportionment

In this approach, the costs of each project activity are calculated separately, and a separate cost-effectiveness ratio is calculated for the outcome of each activity (Dhaliwal *et al.* 2012). For example, in the case of an intervention providing cash transfers and health visits, the costs incurred in each activity are separately calculated and then divided, respectively, by the change in school attendance and morbidity. This approach is very appealing because of its conceptual and computational simplicity. However, as already noted, it can only be applied in the special case in which the different project activities are different interventions. A precondition for this approach to work is that there is a one-to-one mapping between costs $(c_1, c_2, ..., c_n)$ and outcomes $(b_1, b_2, ..., b_n)$ and there are no interactions between the activities. A further requirement of cost apportionment is that each activity must have only one outcome. Owing to these restrictions, this method does not address the challenges of assessing the cost-effectiveness of complex interventions satisfactorily.

We found that only one of the studies that were reviewed employed this approach, this (partial) exception of Abou-Ali *et al.* (2009). Abou-Ali *et al.* assessed the cost-effectiveness of a social funds intervention in Egypt which consisted of separate interventions in education, health, water and sanitation, roads, and microcredit. The costs of each intervention are separately calculated and cost-effectiveness ratios are calculated for each outcome. For example, the total cost of the social fund was 3 billion LE and the cost of the education component was 200 million LE. The authors use the latter figure to calculate the cost of having one less illiterate person. Similarly, they calculate the cost of saving one life under-five (using the total health costs), the cost of one less person with renal disease (using total water and sanitation costs), and the cost of creating one job (using total road costs). This study exposes some of the difficulties of this approach when it is applied to the assessment of integrated projects. First, cost figures for each intervention are crude and not available at any level of detail. Second, outcomes of each intervention are likely to be influenced by the other interventions. For example, the number of lives saved is affected by the health intervention but also by the road intervention so that attribution of intervention costs to a single outcome is rather arbitrary. Third, each intervention has several outcomes and it is not realistic to assign all cost to a single outcome.

4.1.3 Cost–utility analysis

The cost–utility approach explicitly addresses multiple outcomes by aggregating utilities produced by the outcomes. The approach is based on the estimation of utility (U) through a utility function specified for b outcomes (b_1, b_2, ..., b_n):

$$U=f\ (u_i(b_i))$$

The utility so obtained is then used to calculate a cost (C)–utility ratio:

$$CUR=C/U$$

The estimation of an overall utility of the intervention assumes knowledge of the utilities associated to each outcome and of the functional form used for their aggregation. This is not a simple task. Quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) used by health economists are applications of this approach. QALYs and DALYs aggregate all outcomes in terms of life years gained weighted by the quality of living under different levels of morbidity and disability. For example, in QALY, weights represent preferences over different health states, and are obtained through hypothetical lotteries conducted with experimental samples of subjects over different health states and are obtained through hypothetical lotteries conducted with experimental samples of subjects. The characteristics employed in developing the weights include such aspects of life as cognitive skills, physical strength, and emotional wellbeing, which are all crucial in the development of human capital. It is not obvious, however, how similar indices could be calculated in other sectors such as education or infrastructure or how an overall utility index could be calculated for all outcomes of all sectors. On the other hand, the problem of including other economic benefits and costs of health interventions has been acknowledged in health economics (Drummond *et al.* 2005). Health interventions can inflict costs to project beneficiaries; for example, by using their working time (c_e), as well as benefits (b_e), for example by increasing their productivity. These benefits and costs can be calculated and aggregated to project costs in the calculation of the cost–utility ratios, though this is rarely done:

$$CUR=(C+b_e-c_e)/U$$

We did not include cost-effectiveness analyses of health interventions using QALYs and DALYs in our review unless they explicitly tried to account for non-health outcomes of the intervention in this way.

We found only one cost–utility analysis that considered non-health benefits (Miller *et al.* 2013). Miller *et al*. assessed the cost-effectiveness of an HIV prevention intervention, which provided school support to orphan girls in Zimbabwe. They estimated the impact of the intervention through an RCT on three outcomes: early marriage, years of schooling, and health-related quality of life. In order to include the non-health outcomes in the cost–utility analysis, they estimated the returns to schooling (wages) resulting from increased years of education,

and the savings in medical costs resulting from the reduction in early marriage and therefore in HIV infection. They then subtracted these non-health benefits from project costs. The authors conclude that the intervention is highly cost-effective in comparison to the World Health Organization (WHO) benchmark. This approach is applicable to cases in which all but one of the outcomes can be expressed in monetary terms, and when the outcome that cannot be expressed in monetary terms is the relevant one. This study too, however, has some limitations. First, it is not clear that the evaluation included all the relevant outcomes of the intervention such as, for example, learning outcomes. Second, some of the preferences in relation to productivity gains and marriage choices might be already incorporated in the QALY weights, thus leading to a double counting of project effects. Finally, several of the cost calculations and estimations of benefits were rather *ad hoc* and arbitrary, which might be inevitable when trying to monetise all outcomes of an intervention.

4.1.4 Cost–benefit analysis

Another approach to aggregate monetary and non-monetary benefits consists of expressing all non-monetary outcomes in monetary terms using opportunity costs and shadow prices. Cost–benefit analysis compares the streams of all project benefits ($B=b_1+b_2+\ldots+b_n$) and all project costs ($C=c_1+c_2+\ldots+c_n$), all expressed in monetary terms and discounted over time t at the rate r. One typical measure is the net benefit ratio (*NBR*):

$$NBR = \sum_{t=1}^{n} \frac{B_t - C_t}{(1+r)^{t-1}}$$

The *NBR* allows the economic evaluation of any project, not just in comparison to other projects, but also in absolute terms. It is able to tell whether a project is preferable to others and if a project is worthwhile regardless of other projects. The main limitation of this approach is that rarely all benefits can be expressed in monetary terms, unless the researchers are willing to make strong assumptions and several questionable imputations.

We found two cost–benefit analyses in our review (Baird *et al.* 2012; Bernal and Fernández 2013). Bernal and Fernández (2013) assessed the cost-effectiveness of an early childhood intervention in Colombia. They estimate the impact of the programme by length of beneficiary exposure using propensity score matching on the following outcomes: nutritional status, four indicators of socioemotional skills, and six indicators of cognitive development. The positive impact in each domain is translated into wage gains using the results found in the impact evaluation literature. The authors calculate the different values of the cost–benefit ratio depending on the child category considered and using different discount rates. Though the programme effects are, in general, modest, they found the programme to be cost-effective for children exposed for longer than 15 months. The limitations of this study are the following. First, the wage gains are estimated by applying parameters calculated by studies using data from different populations

from the one analysed and no sensitivity analysis is performed. Second, the different effects of the programme on wages are simply added to each other, thus assuming they are independent. However, if, for example, both nutritional and emotional improvements contribute to cognitive development directly and indirectly by affecting each other, this procedure results in double counting. Third, it is not obvious that the three selected indicators represent the whole impact of the intervention and other effects might be present. Finally, no allowance is made for the uncertainty of impact estimates due to sampling variation.

Baird *et al.* (2012) build on earlier work by Miguel and Kremer (2004) already discussed above. The authors revisit a sample of individuals ten years after the implementation of a randomised deworming intervention. They estimate the long-term impact of the intervention on education, employment, labour supply, and productivity, applying difference-in-differences analysis to the original project and control groups. They estimate the wage gains determined by an increase in hours worked. They propose assessing cost-effectiveness employing a *welfare* approach and a *social-planner* approach. In the first case, they compare wage gains to subsidy costs borne by the government and find that the gains largely exceed costs, and that government tax revenues generated by the programme largely compensate for the subsidy. Using the second approach, they calculate the internal rate of return of the project using discounted streams of earning gains and subsidy costs. The returns are shown to be four times the current interest rate, again showing the effectiveness of the project. This study, together with the earlier study by Miguel and Kremer (2004) is probably the best attempt to evaluate the cost-effectiveness of a complex intervention that we were able to find in our review. It does, however, have some limitations. First, no allowance is made for uncertainty of the results because of sampling variation. Second, cost estimates are based on a programme implemented in a different area and population.

5 Conclusions

Our review was able to find only seven studies assessing the cost-effectiveness of complex interventions. This is certainly a reflection of the limited use of cost-effectiveness analysis in the practice of impact evaluation. Despite the surge of impact evaluation studies in recent years, few cost-effectiveness analyses are conducted alongside impact evaluations. It should be emphasised, however, that the small number of studies found is also the result of the difficulty of identifying cost-effectiveness studies and of the restrictive selection criteria adopted. We identified studies by screening titles and abstracts, but cost-effectiveness analyses are often conducted as a secondary goal by impact evaluations and may not be reported in the title and abstract. In addition, we limited the review to impact evaluations using experimental and quasi-experimental designs, and did not consider cost-effectiveness analysis employing the results of other impact evaluations, of which there might be many. Finally, we defined complex interventions as interventions with multiple outcomes across sectors. This was done

mainly with the goal of excluding all cost–utility analyses using QALYs and DALYs produced by health economists, but which may have resulted in the exclusion of relevant studies in other sectors as well.

All the studies reviewed concluded that the interventions were cost-effective. These conclusions, however, are tempered by the methodological problems involved in assessing the cost-effectiveness of complex interventions. We identified the use of four different methodologies: cost–consequence analysis; cost apportionment; cost–utility analysis; and cost–benefit analysis. Each of these approaches may be employed effectively in specific cases, but none can be applied in all cases. Cost–consequence analysis is simple and easy to use, but it requires a cost-effectiveness comparison between few projects and few outcomes. Cost apportionment is a straightforward method of assessing cost-effectiveness, but it requires that each project component has a single outcome and that project components are independent. Cost–utility analysis has been applied very successfully in the health sector, but it is unclear whether utility indices like QALYs and DALYs can be developed in other sectors such as, for example, education and governance, and it is even less clear whether a single utility index can be formulated for all outcomes across all sectors. Finally, cost–benefit analysis effectively assesses the welfare impact of an intervention, but not all outcomes can be monetised unless we are willing to accept some peculiar assumptions.

In addition to the methodological difficulties outlined above, the studies reviewed shared a few other limitations. First, they rarely considered all the intended and unintended outcomes of the interventions. The choice of the outcomes often appeared to be motivated more by the availability of data rather than by a solid theory of how the interventions determine the outcomes. Second, none of the studies reported confidence intervals of the cost-effectiveness ratios. A meta-analysis of cost-effectiveness ratios of primary education interventions by McEwan (2014) shows how the inclusion of confidence intervals may considerably change the policy conclusions of a cost-effectiveness study. Finally, all studies suffered the practical difficulties of obtaining cost data and only two studies included social costs.

In summary, our review found few cost-effectiveness studies of complex interventions, no widely applicable methodologies, and a number of practical problems in measuring the costs and effects of the interventions. Much could be improved by conducting more cost-effectiveness studies along impact evaluations, and by exercising more care in the calculation and reporting of costs and outcomes. However, 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 the cost-effectiveness of complex interventions. We praise the studies included in this review for making the effort to assess cost-effectiveness across a multiplicity of outcomes. Most development interventions are complex, and the first wrong assumption made by many cost-effectiveness studies is that they are not.

Annexe 1 Data extraction sheet

Section 1 Study details	
1.1 Authors	Author names
1.2 Title	Title of the study
1.3 Type of study	Indicate type of study and source, for example: 'Journal article, Journal of Development Effectiveness' or 'Working paper, IFPRI'
1.4 Brief description of programme/ intervention	A brief description of the project evaluated
1.5 Sector	Sectors of intervention such as education, health, etc.
1.6 Country (-ies) of implementation	Countries
1.7 Method used in the impact evaluation	Impact evaluation method: RCT, RDD, PSM, DID
1.8 Outcomes of the programme	Outcomes of the intervention
Section 2 Cost data	
2.1 Government or social	Specify which cost is considered: government (project budget); social (including shadow prices, costs to beneficiaries, costs to society, etc.)
2.2 Ingredients (1) or all budget (2)	Specify which costing method was used: ingredients (all the components of the overall cost), or all budget (the whole budget cost is included in the CEA)
2.3 If the ingredients approach is used, what cost categories are included?	List the cost categories used; for example, direct costs of the transfer, costs for delivering the transfer to the beneficiary, etc.
2.4 Quality assessment	An assessment of the quality of the cost data and whether enough information is given in the paper
2.5 Others	Any other relevant observation
Section 3 Methods for cost-effectiveness analysis	
3.1 Method category (codes 1 to 4)	(1) Total cost of the programme divided by the outcomes, separately for each of the outcomes. (2) The cost of the relevant component divided by the relevant outcome. (3) Welfare approach comparing costs and benefits (CUA or CBA). (4) Any other method.
3.2 Description of method	Brief description of the method used
3.3 Cost apportionment (YES/NO)	Whether cost apportionment was made or not
3.4 Treatment of multiple outcomes	Method to address cost-effectiveness of multiple outcomes
3.5 Discounting (YES/NO)	Whether discounting was used or not
3.6 Do authors flag a concern related to the treatment of multiple outcomes? (YES/NO)	Whether multiple outcomes issue is discussed
3.7 Do the authors mention synergies? (YES/NO)	Whether synergies between activities are discussed
3.8 Do the authors discuss or consider social costs (even if separate from CEA)? (YES/NO)	Whether social costs are considered
3.9 Quality assessment	Overall assessment of the cost data
3.10 Others	Any other relevant observation

Source Authors' own.

Annexe 2 Study details

Table A1 **Details of the studies included in the review**

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Ahmed *et al.* (2009)	Comparing Food and Cash Transfers to the Ultra-Poor in Bangladesh	A World Food Programme (WFP) social protection programme in Bangladesh. Four different interventions targeting poor women: (1) food rations; (2) food and cash; (3) food and cash as wage payments for public works; (4) cash as wages in road construction.	PSM	Food consumption; caloric intake; nutritional status; income; poverty; expenditure of former beneficiaries.	Ingredients approach (direct costs of the transfers and of delivering the transfers). No social costs. No cost apportionment.	The cost of each programme (including the transfer and the delivery cost) is divided by each outcome separately.	A ranking of the programmes is provided in relation to just one outcome (poverty reduction).
Hidrobo *et al.* (2012)	Cash, Food or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador	A WFP social protection programme in Ecuador. Three different transfer modalities: (1) cash transfers; (2) food rations; (3) food vouchers.	RCT	Food expenditure; caloric intake; dietary diversity score, dietary diversity index, food consumption score.	Ingredients approach. After accounting for common costs, the differences in marginal costs are considered. Social costs considered but not included in CEA.	The cost of each programme is divided by each outcome separately.	Food is the most expensive method on all outcomes while vouchers are cheaper than cash transfers on four out of five outcomes.
Miguel and Kremer (2004)	Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities	A deworming and health education intervention by an NGO in a small district of rural Kenya.	RCT	Infection rates by four types of worms; school attendance; maths, English, and science test scores.	It uses the cost of US$0.49 per pupil from a similar project in Tanzania because the programme is not running to scale and because it was difficult to remove the evaluation cost from the project cost. No social costs.	It calculates the cost per DALY for the worm infections; cost per additional years (or days) of schooling; the cost per education returns to schooling. All ratios use the same project cost figure.	The cost per DALY is compared to a benchmark while the other cost ratios are compared to those of other primary education interventions in Kenya. Deworming is found to dominate other projects.

Table A1 Details of the studies included in the review (cont.)

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Abou-Ali *et al.* (2009)	Evaluating the Impact of Egyptian Social Fund for Development Programs	The Egyptian social fund for development established in 1991. It includes (1) a community development programme (literacy classes, primary health care, small environmental projects); (2) a public works programme (potable water, sewage; and roads); (3) a microcredit programme (loans).	PSM	Illiteracy rate, employment rate, immunisation rate, under-five mortality rate, health expenditure, diarrhoea incidence, water-related morbidity, income and poverty.	It uses the total cost of the three programmes over the period 1991–2007. A crude apportionment of costs to project components is performed. No ingredients approach. No social costs.	For sectors, the cost of the relevant component is divided by the relevant outcome. For outcomes within sector, the total cost of the programme is divided by the outcomes, separately for each of the outcomes.	Most projects do not appear to be very cost-effective although to varying extents.
Miller *et al.* (2013)	Cost-Effectiveness of School Support for Orphan Girls to Prevent HIV Infection in Zimbabwe	School support provided to Grade 6 orphan adolescent girls in a province of rural Zimbabwe during 2010–14. Support consisted of a package of fees, uniforms, school supplies, and teacher training.	RCT	Three outcomes: school retention (metric is years of schooling); early marriage (metric is years unmarried); QUALY (metric is a health-related self-assessment on a five-dimensional scale).	It uses the ingredients approach for 11 budget items. No social costs.	Cost per QALY. Multiple outcomes are accounted for by subtracting the monetarisation of other benefits from project costs.	The intervention is found to be highly cost-effective in relation to the WHO QALY threshold.
Baird *et al.* (2012)	Worms at Work: Long-Run Impacts of Child Health Gains	This is the same deworming intervention evaluated by Miguel and Kremer after a ten-year follow-up.	RCT, DID	Self-reported health, hours of work and meals per day.	It appears to use an ingredients approach. Social costs are included by considering the costs of subsidising the intervention.	They summarise the welfare benefits of the programmes in terms of wage gains. They compare wage gains to the cost of subsidy and calculate the internal rate or return of the project.	They find that the intervention is cost-effective.

Table A1 **Details of the studies included in the review (cont.)**

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Bernal and Fernández (2013)	Subsidized Child Care and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure	A home-based early child development programme for children under five, delivering childcare, supplemental nutrition, and psychosocial stimulation.	PSM	Cognitive abilities, nutritional status, and socio-psychological development.	Ingredients approach including parents' co-payments, but not other costs incurred by families.	All impacts are translated into wage increases using the results from other studies. A benefit–cost ratio is calculated.	The programme may increase future earnings though the benefit–cost ratio is modest.

Source Authors' own.

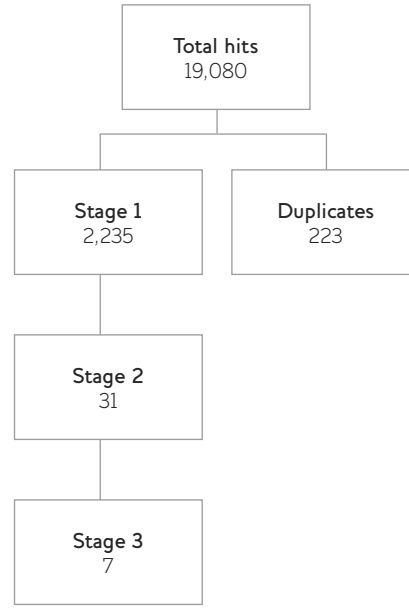
Annexe 3 Data extraction and search results

The studies finally selected for the review were independently analysed by three reviewers, covering questions regarding costing and cost-effectiveness methods (see Annexe 1). In particular, the methods employed to collect cost data were reviewed and the quality of such data assessed when possible. Further, the reviewers extracted the cost-effectiveness method applied and how it accounted for the presence of multiple outcomes (see Table A1 in Annexe 2).

The results of the search and selection process are illustrated in Figure 2. Search of the databases returned a total of 19,080 hits. Removal of duplicates and a first screening based on titles and abstract, which removed studies not relating to low- and middle-income countries and studies not conducting cost-effectiveness analysis, led to the selection of 2,235 studies. Further review of titles and abstract and a sequential application of the following selection criteria: (1) low- and middle-income country; (2) complex interventions; (3) cost-effectiveness analysis; (4) impact evaluations; and (5) analysis of multiple outcomes, led to the selection of 31 studies. Finally, an in-depth review of the 31 studies selected led to the exclusion of 24 other studies as closer inspection revealed they did not conform to the five selection criteria above. Only seven studies were selected for the final review.

Figure A1 Flow chart of the selection process

Medline: 3,527; ERIC: 270; Web of Science: 9,056; EconIlt: 432; Ideas: 4,434; J-Pal: 160: World Bank: 878; IFPRI: 200; 3ie: 123; Cochrane: 51; DFID: 97; Campbell: 247; EPPI: 199



Stage 1: By reviewing titles and abstracts we excluded: studies from high-income countries; non-cost-effectiveness studies; and duplicates

Stage 2: By reviewing titles and abstracts we sequentially selected: studies on low- and middle-income countries (1,920); complex programmes (992); cost-effectivenss analyses (70); impact evaluations (31)

Stage 3: In-depth review led to the exclusion of a further 24 studies

Source Authors' own.

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 Centre of Excellence for Development Impact and Learning (CEDIL) at the London School of Hygiene & Tropical Medicine.

2 Institute of Development Studies, Brighton, UK.

3 Honorary Assoc. Professor, London School of Hygiene & Tropical Medicine (2012–17).

4 International Fund for Agricultural Development (IFAD), Rome, Italy.

5 Institute of Development Studies, Brighton, UK.

6 The PICO model is widely used in the synthesis of evidence as a strategy to formulating questions and organising a literature search. PICO stands for Population or Problem (what are the characteristics of the project population or the nature of the problem considered?), Intervention (what is the intervention?), Comparison (what is the counterfactual?), and Outcomes (what are the relevant outcomes?).

7 This systematic review has not been registered with the Campbell Collaboration due to the difficulties of matching the criteria for a review *of methodologies* like this one, rather than the impact evaluation *of results*. However, we followed closely the Campbell Collaboration Systematic Review Guidelines. Due to the nature of our research question, we made no attempt to summarise the evidence in a quantitative way.

References

Abou-Ali, H.; El-Azony, H.; El-Laithy, H.; Haughton, J. and Khandker, S.R. (2009) *Evaluating the Impact of Egyptian Social Fund for Development Programs*, World Bank Policy Research Working Paper 4993, Washington DC: World Bank

Ahmed, A.U.; Quisumbing, A.R.; Nasreen, M.; Hoddinott, J.F. and Bryan, E. (2009) *Comparing Food and Cash Transfers to the Ultra-Poor in Bangladesh*, IFPRI Research Monograph 163, Washington DC: International Food Policy Research Institute

Baird, S.; Hicks, J.H.; Kremer, M. and Miguel, E. (2012) 'Worms at Work: Long-Run Impacts of Child Health Gains', unpublished paper

Bathia, M.R.; Fox-Rushby, J. and Mills, A. (2004) 'Cost-Effectiveness of Malaria Control Interventions when Malaria Mortality is Low: Insecticide-Treated Nets Versus In-House Residual Spraying in India', *Social Science and Medicine* 59.3: 525–39

Bernal, R. and Fernández, C. (2013) 'Subsidized Child Care and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure', *Social Science and Medicine* 97: 241–49

Dhaliwal, I.; Duflo, E.; Glennerster, R. and Tulloch, C. (2012) *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education*, Cambridge MA: Abdul Latif Jameel Poverty Action Lab (J-PAL), Massachusetts Institute of Technology (MIT)

Drummond, M.F.; Sculpher, M.J.; Torrance, G.W.; O'Brien, B.J. and Stoddart, G.L. (2005) *Methods for the Economic Evaluation of Health Care Programmes*, 3rd ed., New York NY: Oxford University Press

Hidrobo, M.; Hoddinott, J.; Peterman, A.; Margolies, A. and Moreira, V. (2012) *Cash, Food, or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador*, IFPRI Discussion Paper 01234, Washington DC: International Food Policy Research Institute

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

Mauskopf, J.; Paul, J.E.; Grant, D.M. and Stergachis, A. (1998) 'The Role of Cost–Consequence Analysis in Healthcare Decision-Making', *PharmacoEconomics* 13.3: 277–88

McEwan, P.J. (2014) 'Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomised Experiments', unpublished paper

Miguel, E. and Kremer, M. (2004) 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', *Econometrica* 72.1: 159–217

Miller, T.; Hallfors, D.; Cho, H.; Luseno, W. and Waehrer, G. (2013) 'Cost-Effectiveness of School Support for Orphan Girls to Prevent HIV Infection in Zimbabwe', *Prevention Science* 14.5: 503–12

Sabet, S.M. and Brown, A.N. (2018) 'Is Impact Evaluation Still On the Rise? The New Trends in 2010–2015', *Journal of Development Effectiveness* 10.3: 291–304

This page is intentionally left blank

Assessing Value for Money in Integrated Development Programmes – The Case of a Millennium Villages Project in Northern Ghana*

Arnab Acharya[1] and Tom Hilton[2]

Abstract Through the use of cost–consequence analysis (CCA), a recent evaluation of a Millennium Villages Project (MVP) in Ghana revealed it to have represented poor value for money (VFM), with comparator projects elsewhere seeming to deliver similar results at less than half the cost. However, complex integrated development programmes (IDPs) such as the MVP pose serious challenges for VFM assessments. IDPs target system-wide changes in resource-scarce contexts, making expensive foundational investments in infrastructure and other systems. The unit costs of benefits will tend to be high in the short or medium term. In contrast, many standalone projects, showing greater efficiency, may target similar outcomes, but do so by building upon existing prior investments. In this article, comparing three VFM approaches, we argue that CCA is the most appropriate for IDPs. However, its applications must be mindful of the contextual differences in which the comparator standalone projects and the IDP were implemented.

Keywords: evaluation, value for money, cost–benefit analysis, cost-effectiveness analysis, cost–consequence analysis, evaluability, integrated development programme, Millennium Villages Project.

1 Introduction

International development programming has become increasingly complex in recent years, particularly because a global agenda for social, economic, and environmental change has been shaped by the Millennium Development Goals (MDGs) and the subsequent Sustainable Development Goals (SDGs). The holistic approaches embodied by such 'integrated development programmes' (IDPs) pose significant challenges for evaluators. Greater complexity of programming makes it difficult to (a) untangle multiple strands of

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.161
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'**; the Introduction is also recommended reading.

intervention, (b) agree upon standardised indicators of success, and (c) identify suitable comparators for evaluations. In this article, we use the example of a recent evaluation of a Millennium Villages Project (MVP) in the Savannah Accelerated Development Authority (SADA) region of northern Ghana to illustrate these challenges.

The article proceeds as follows. Section 1 sets out the theory behind IDPs such as the MVP, before Section 2 presents various approaches to measuring value for money (VFM), and the rationale for the choice of the VFM approach for the MVP in northern Ghana. Section 3 presents a summary of the results of this VFM analysis,[3] and Section 4 provides a discussion of the difficulties in interpreting VFM results in the context of an IDP. Section 5 summarises our conclusions.

1.1 Integrated development programming

A dominant early view in development economics was that of the 'big push', whereby substantial investment across productive sectors would allow countries to break free from the poverty traps and move towards an accelerated growth process.

Subsequent theories of economic growth, known collectively as 'endogenous growth theory', promoted the view that economic growth paths not only depend on physical capital accumulation, but also on health and education – human capital, which in turn prompt innovations in technology, management, and finance (see Romer 2006). Along this line, Gallup and Sachs (2001) provided econometric evidence to show that there is a directional causality between the burden of illness and the performance of the economy in many developing countries.

The policy implication emerging from these theories is that, within a politically stable environment, IDPs emphasising technological and business innovations, along with educational, health, and infrastructure improvements is the key to breaking poverty traps and stimulating development.

This approach complemented the emerging consensus that gross domestic product (GDP) alone is a poor measure of wellbeing, and that broader measures of 'human development' (including health and education) were also needed (UNDP 1990). This view found its origin in Amartya Sen's human capability approach which argued that access to income did not necessarily provide many in the world with an adequate quality of life (Sen 1985).

In light of this, the MDGs established a set of multidimensional global targets that would come to shape the global development agenda through the 2000s, looking not only at income levels, but also health, education, and an array of other metrics. This multidimensional philosophy led to the rise of complex IDPs which sought to deliver a wide range of results within a single programme.

Figure 1 VFM measures

Degree of subjectivity	**VFM measure**	**Comparators**	
Low ↓ High	**Cost–benefit analysis** All benefits monetised. Analysis presented as a ration of benefits-to-costs (i.e. '2.4:1'). Easy to compare standardised measure across different investments.	Comparison with benefit-to-cost ratios of other projects. (However, a lack of availability of pure CBA figures means projects are often assessed on whether their own benefits outweigh their costs.)	$ vs. $
	Cost-effectiveness analysis Benefits not monetised but well-defined and comparable with other projects (i.e. 'years of schooling'). Analysis presented as cost-per-output figures, for well-defined outputs with comparators.	Comparison with costs per output values (measurable by the output itself or the benefits produced by the output) of other projects targeting benefits measurable by a common indicator.	$ vs. Output
	Cost–consequence analysis Benefits not monetised. Range of benefits, often resulting from a complex project, compared with overall project cost. Difficulty of finding appropriate comparators makes value judgements quite subjective.	Hard-to-find projects targeting the same range of outputs. Often compared instead with the sum of a range of individual cost-per-output values from other projects.	$ vs. Output 1 Output 2 Output 3 Output 4

Source Authors' own.

1.2 The MVP evaluation

The Millennium Villages Project of the Earth Institute at Columbia University is one example of this kind of programming. Starting in 2004, the MVP applied an IDP approach, focusing on education, health, agriculture, and infrastructure, at 14 sites across ten sub-Saharan African countries. Since its inception, the MVP has attracted considerable criticism relating to both the high cost of implementation and the limited evidence on results generated.

The difficulty in evaluating the MVP has also been noted, given the lack of comparable projects and non-experimental project design (see, for example, Tollefson 2015). This article uses the results from the evaluation (Barnett *et al.* 2018) of one MVP site, in the SADA area of northern Ghana, to further explore some of these challenges, and specifically to show how definitive judgements relating to value for money are difficult for IDPs.

2 Choice of VFM measure

We begin with a clarification of what is meant by VFM, and a summary of a number of approaches that are used to measure it.

VFM assessments seek to measure the extent to which a particular investment or project represented the best possible use of the funds in the pursuit of a particular goal. A range of VFM approaches can be applied depending on the availability of data relating to the costs and benefits of the project, as well as the availability of comparators (see Figure 1). In each case, the costs of a project are weighed up against the benefits that these costs generate.[4]

2.1 Cost–benefit analysis

Cost–benefit analysis (CBA) monetises all costs and benefits to give an overall benefit-to-cost ratio – a headline figure that clearly expresses whether or not benefits have outweighed costs on a particular project, and which allows for comparison with other potential investments and their respective benefit-to-cost ratio. CBA requires all benefits to be monetised, and as such, is often limited to projects where targeted returns are largely financial in nature.

2.2 Cost-effectiveness analysis

Where benefits cannot be readily monetised, cost-effectiveness analysis (CEA) is more commonly used. CEA provides an approximate unit cost for producing a particular desired output. Unlike CBA, its applicability requires comparator programmes that are aiming to produce the same outputs, allowing for comparisons of, for example, 'cost per year of schooling' under a range of different delivery models.

CEA has commonly been used in the health sector – whilst health outputs are often not easy to monetise (in order to conduct CBA), they are often relatively easy to define and compare across interventions – i.e. cost per vaccination, or cost per mosquito net distributed. There is agreement among researchers in the field of health policy that health benefits or health utility can be measured through single units of measure such as quality-adjusted life years (QALYs) (see Tan-Torres Edejer *et al.* 2003). Thus, the terms cost–utility analysis and cost-effectiveness analysis are commonly used to assess programmes that produce varied types of health benefits but yield to a single common unit of measure.

2.3 Cost–consequence analysis

With the rise of IDPs, projects often seek to influence multiple aspects of wellbeing which cannot easily be summarised with a single benefit measure.[5] In such cases, evaluators may choose to list the various individual outcomes alongside the costs – this is known as cost–consequence analysis.

Within cost–consequence analysis, there are two possible approaches:

- **Cost-apportionment:** One approach spreads the total programme costs across multiple activities that produce specific outcomes, and then determines the unit costs of each outcome, which can be compared to unit costs for producing these outcomes through alternative means (as per CEA, above).

  However, cost-apportionment may not always be easy – there may be sizeable portions of costs stemming from common activities (such as programme management and overheads) that produce multiple outcomes; moreover, any one outcome could be produced through the partial efforts of several types of activities as the results of multiple simultaneous interventions reinforcing one another.

- **Total cost:** Another approach is to assess whether or not the total cost of providing a range of benefits $\{b_1, \ldots, b_n\}$ under a single IDP (C_{Joint}) is lower than the summed costs of multiple separate projects delivering the same results at the unit costs of $\{c_1, \ldots, c_n\}$:

  $$C_{Joint}\, b_1, \ldots, b_n < \sum_{i=1}^{n} c_i \text{ x } b_i$$

  The inequality shows costs savings or efficiency due to synergistic interventions within IDPs, as individual activities within an IDP share programme overheads and reinforce each other's results. As with CEA, CCA still poses challenges to finding appropriate comparators, which may be required for a wide range of outputs.

Given the complexity of the programme, CCA was the VFM measure of choice for the MVP, using a combination of the cost-apportionment and total cost approaches outlined above. Section 3 sets out the results of this analysis.

3 MVP cost–consequence analysis

This section presents a summary account of the observed costs and benefits of the MVP in northern Ghana explained in more detail elsewhere (Barnett *et al.* 2018). First, cost estimations are offered through the use of standard methods for accounting public projects (see Tan-Torres Edejer *et al.* 2003) for the years 2012 to 2016 mid-year, the time at which the project impacts were last measured. We then provide measures of benefits, followed by the combined cost–consequence analysis.

Costs and benefits were categorised under six primary sectors: health, education, agriculture, infrastructure, community development, and environment.

3.1 Costs

Mitchell *et al.* (2018) report spending for an array of previous MVP projects in sub-Saharan Africa. They indicate that the annual per capita costs ranged from US$109 to US$132 in 2005 US prices. This figure is comparable to what we can report for the MVP in northern Ghana, which at current dollars is US$123, valued at 2014 prices. The total expenditure for the MVP in northern Ghana for 2012–16 amounted to US$16 million (current US$). This is the allocated amount or budget for the project; in reporting costs, this figure needs to be adjusted to account for capital goods purchased that were used across all the years of the project and some which can be used beyond the project.

The four largest sectors in terms of expenditure were health, infrastructure, education, and agriculture. In order to calculate the economic cost of the project, we adjust the allocated budget the following way:

- First, overheads not linked to any specific programme area, totalling US$4.7 million (including management, administration and operation, monitoring and evaluation, and technical assistance), were distributed proportionately across the six primary sectors.

Table 1 **Per capita annual costs by sector for the years 2012 to mid-year 2016 (all US$, 2012 net present value)**

	Health	Education	Agriculture	Environment	Infrastructure	Community development	Total	Annual
All Cost[1]	156	61	71	5	71	31	395	88
Limited Mgmt[2]	146	57	66	5	66	29	370	82

Notes (1) These figures include cost of management, operating and administration, technical assistance, and M&E.
(2) These figures include only cost of management, operating, and administration costs.
Source Barnett *et al.* (2018).

- Second, costs were distributed according to their usage years. Many of these purchases had immediate usage and were only counted in their purchase year. Other goods (capital goods) had usage beyond the year they were purchased and were distributed across multiple years accordingly. The value of these capital goods was obtained through annuitisation of costs using a 5 per cent discount rate.
- Third, costs were only counted to mid-2016, so as to ensure comparability with a mid-2016 measure of benefits. Annuitised costs falling beyond mid-2016 were treated as 'investment' costs, for which

Table 2 **Cost–consequence analysis of the MVP**

MVP indicators	Observed results	Estimated MVP costs[9]	Comparator costs
Income	Total: US$3.8m	US$1.8m to US$3.6m	US$2.8m[10]
Net attributable income gain	(per capita: US$141; per household: US$1,001)		
Education	2,153 school years (valued at US$3.1m–US$4.3m)	US$1.6m	US$0.2m[11]
Net primary school attendance			
Health		Total package:	Total package:
Malaria reduction	891 fewer cases of malaria	US$4.2m	US$0.4m to US$0.6m[12]
Stunting	467 fewer cases of stunting		
Health-care worker contacts	48,500 additional contacts		
Vaccination	832 vaccinations (all types)		
Contraception	2,112 couple year of protection		
Sanitation	8,187 people affected per year	US$0.4m	(no relevant comparator available)
Total costs		US$10m	US$4m to US$5m

Source Barnett *et al.* (2018).

future benefit streams were expected to arise. As of mid-2016, the project had invested US$2.8 million to be used after 2016, including US$1.4 million in infrastructure.[6]

- Finally, all costs were discounted at a rate of 5 per cent to arrive at 2012 US$.

Accounting for these factors, we estimate that the total costs associated with the benefits observed to date to be US$10 million.

These costs were split across the average annual population of the project area to arrive at the per capita costs shown in Table 1. Two scenarios are presented – one including all management and overhead costs ('All Cost'), and one including only the essential management functions that might be included should the programme be replicated (removing substantial technical support and monitoring and evaluation (M&E)) ('Limited Mgmt.').

3.2 Benefits

To understand the benefits of the MVP, we consider the different rates of change in indicators of interest in both the Millennium Village sites (MVs) and the control villages (CVs), in order to estimate the additional impact attributable to the project at the project site compared to the counterfactual of no intervention in the MV area (MV'). This 'difference-in-differences' (DiD) approach is summarised as follows:

$$(MV_{2016} - MV_{2012}) - (CV_{2016} - CV_{2012}) = (MV_{2016} - MV_{2012})\,(MV'_{2016} - MV'_{2012})$$

DiD assumes that, in the absence of the project, the difference in indicators at both sites would have been the same due to parallel trends in both areas. As outcomes in the MV areas are the same at the beginning of the project with or without the project in 2012, the equation reduces to the following statement of additionality versus the counterfactual by the end of the project period:

$$DiD = MV_{2016} - MV'_{2016}$$

The MVP essentially had three major goals:

1 To raise income through improved agricultural productivity, market development, micro-finance, and better infrastructure.

2 To improve health through the development of a better health system, delivering primary care at village level, with a referral system that links up with Ghana's newly developed national health insurance scheme, promoting universal health coverage.

3 To improve educational achievements through improving the educational system.

For each of the above, the DiD approach was used to estimate the net attributable benefits in the MVs. A summary of these results is set out in Table 2.

3.3 Cost–consequence analysis

A summary of the cost–consequence analysis for the MVP in northern Ghana is presented in Table 2, including the observed benefits of the project, the estimated MVP costs associated with these benefits, and a range of comparator costs drawn from the literature.[7]

With regard to MVP benefits, income benefits were derived from a combination of agriculture, infrastructure, and community development, totalling some US$3.8 million. Education benefits in terms of additional years of schooling were monetised to arrive at an estimated value of US$3.1 million to US$4.3 million (Fink *et al.* 2016; Montenegro and Patrinos 2014).[8] We did not observe a health impact that could be measured by health utility level, perhaps due to the duration being too short between the observation period and the time at which much of the health infrastructure was placed. Health benefits are assessed at the output level, including access to health-care workers, vaccinations, stunting, contraception, and malaria cases avoided.

The MVP costs are listed as 'estimates', since the MVP cost data were not mapped to specific activity areas within sectors, and there is a degree of crossover between sectors. Furthermore, we are only able to provide aggregated costs at the sectoral level – for example, it is impossible to tell from the data exactly how much was spent on malaria reduction. Instead, an overall estimated cost of providing the full package of health benefits is provided.

With regard to comparators, we carried out an extensive literature search to find the cost-effectiveness of producing the same benefits that the MVP produced by other means. The result of our effort has been detailed in Acharya, Masset and Saha (2017). As reported in Masset *et al.* (this *IDS Bulletin*), it was not possible to find any single programme taking a similarly comprehensive approach as the MVP. Instead, individual programmes delivering similar results at the sectoral and sub-sectoral level were sought. Some further limitations apply to the comparators (as is often the case with CCA or even CEA), notably the difficulty in finding (a) projects seeking to deliver the exact same outputs, (b) projects operating in the same locality or similar contexts, and (c) projects using the same costing methods.

Whilst noting these limitations, the literature suggests that the overall benefits generated by the MVP in northern Ghana could have been generated at around half the cost by other means. This suggests that the intensive IDP approach of the MVP has not, overall, represented good value for money.

At the sectoral level, income generation likely had positive returns, and appears to have been delivered in a relatively cost-effective manner. Educational benefits significantly outweigh MVP spending on education. However, examples from the literature suggest that such benefits could have been delivered at far lower cost. Similarly, health

outputs could have been delivered at a fraction of the cost by other means. At the sub-sectoral level, since we could not apportion costs to specific results, we cannot make a judgement as to the cost-effectiveness of, for example, avoidance of malaria cases.

Whilst the overall picture suggests poor value for money, there are some challenges in interpreting CCA results for IDPs. This is discussed in the following section.

4 Interpreting the cost–consequence analysis

Whilst the MVP in northern Ghana appears to represent poor VFM under the CCA calculations, interpretation is just as important as calculation in drawing conclusions about the programme. Careful consideration is needed of the specific context of the MVP in northern Ghana (and IDPs in general), the appropriateness of comparisons with standalone programmes, and the relevant time horizons of the evaluation.

4.1 Contextualising the costs of the MVP

IDPs typically take place in resource-scarce regions characterised by significantly under-developed infrastructure and systems – as such, 'big push'-style IDPs tend to require more intensive investment than may be required in a more developed context, opting to take on the role typically played by governments in making foundational investments.

Indeed, the SADA region is one of the least developed regions of Ghana. To put the MVP budget of US$88 per capita per annum in context, the expenditure of the Government of Ghana amounted to 17.7 per cent of GDP in 2016, or US$268 per capita (World Bank 2018). MVP spending has therefore amounted to the equivalent of around a third of government expenditure over the period, a very substantial sum. However, observations of public services in the SADA region suggest that government spending is likely to have been below the national average in the MV and CV areas. Thus, increasing the budget by US$88 may not be a great burden for an area where much expenditure has not already taken place.

Clemens (2012) noted that the MVP is an expensive programme. Yet, the average per capita for the MVP is not high compared to an important programme that sought to improve livelihoods through providing productive assets to poor households in Ghana (Banerjee *et al.* 2015). The project reported a cost of US$75 per capita annually over three years. The comparable value for the MVP amounted to US$82 per capita annually.

The comparison with standalone projects showed the MVP to be inefficient. Standalone projects targeting specific outcomes, however, typically operate within pre-existing systems, building on existing infrastructure and services. The costs of these foundational elements are not factored into the cost-effectiveness calculations of most programmes. In the absence of reliable cost-effectiveness data on

delivery models within the specific SADA region, the efforts of the MVP may be undervalued, given the challenging context it was operating in.

For example, cash transfer programmes are built on the assumption that health and educational goods and services can be purchased through the payments made to households. The Ghana Livelihood Empowerment Against Poverty (LEAP) programme provides about US$18 per capita per annum in cash transfers (de Groot 2016).[13] However, in the SADA region it is unlikely that the desired health or education services could have been obtained with such cash transfers since many health centres and schools were found to be near non-operational at the outset of the MVP. Additional investments are required in order to build up infrastructure and services before a programme like LEAP can be effective. It may be considered that some of the high costs of the MVP can be attributed to developing these underlying systems, which are not typically valued in the VFM analysis of standalone programmes.

4.2 Evaluating systems-level interventions

Although much has been written regarding interventions at the systems level in health and education, there has been little attempt to develop a means of evaluating systems-level changes (see Hanson 2015).

One motivation for systems-level interventions in IDPs is the anticipation of significant synergistic benefits as programme elements become self-reinforcing, leading to programming that is greater than the sum of its parts, and ultimately more cost-effective than standalone delivery through multiple programmes. For example, a programme to provide uniforms and textbooks together may represent greater value for money than two standalone programmes for uniforms and textbooks, due to a combination of shared overheads and enhanced benefits.

However, such synergies have not yet been observed in the MVP in northern Ghana. One important factor may be the time-lag in impacts. In an evaluation of the development of education systems in Indonesia, Duflo (2001) explores the impact of increasing the size of the school system through the construction of new schools and the development of supporting educational systems. Whilst much of the initial investments occurred in the 1970s, Duflo notes that the impact on wages is identified only 20 years later. Similarly, whilst infrastructure investments can bring about substantial systems changes, these changes have usually been observed after a considerable time-lag (Banerjee, Duflo and Qian 2012).

In the case of the MVP, there may yet be gains in terms of synergistic benefits that are only observed after ten or more years, as opposed to the short-term time horizons of narrower standalone programmes. For example, the health outputs observed in the MVP in northern Ghana to date have been delivered as part of a package of intensive health systems development. This may have rendered the specific health outputs generated significantly more expensive than those of standalone programmes that may have been operating within more developed

health systems. The health outcomes associated with these outputs, such as longer life expectancy and increased productivity, may not be apparent for a number of years. As these improved outcomes take hold, however, they may serve to reinforce the benefits seen elsewhere under the MVP, as healthier people are better able to take advantage of improved agriculture and infrastructure systems.

5 Conclusion

The rise of complex IDPs targeting multiple objectives under the MDGs and subsequent SDGs has presented a number of challenges for evaluators. The difficulty of monetising benefits or otherwise capturing them under standardised all-encompassing indicators, combined with the difficulty in finding comparator programmes, tends to render conventional VFM assessments, such as CBA and CEA, impractical. CCA is one alternative, setting out an array of benefits against the costs of the programme, and seeking to compare this with alternative delivery models. Given the unique nature of IDPs, single comparator programmes are usually unavailable, and comparisons instead tend to be made up of multiple standalone projects.

This creates challenges for the interpretation of results, however. Standalone comparator programmes are often based on the workings of functioning systems needing only marginal changes at relatively low cost – the cost of developing the underlying systems and infrastructure is not typically accounted for. IDPs such as the MVP, on the other hand, may have far greater up-front costs as they focus on building up systems from a lower level of development. The benefits from these investments – particularly the synergistic benefits targeted under broad, systems-development investments – may take many years or even decades to become apparent.

The CCA suggests that the benefits delivered to date under the MVP in northern Ghana could have been delivered at half the cost using alternative means, and few signs of synergistic benefits have yet been apparent. Such a large gap indicates that the MVP failed to achieve value for money. Nonetheless, it is unclear whether the approaches identified in comparator programmes would have been effective in the resource-scarce SADA context, whilst the full benefits of the MVP investments may take several years to become apparent.

In light of the above, future research should focus on (a) revisiting specific IDPs such as the MVP, in order to assess the extent to which investments made with regard to systems and infrastructure are sustained and deliver additional benefits that improve the VFM proposed for the initial investments; and (b) developing new methods to evaluate systems-level investments with high up-front costs and long lags in benefit streams.

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 Assoc. Professor, London School of Hygiene & Tropical Medicine (2012–17).

2 Freelance development economist.

3 Full details of the VFM analysis of the northern Ghana MVP are available in the final evaluation report (Barnett *et al*. 2018).

4 For each case described here, we assume that all costs can be monetised. Whilst this may not always be the case, a discussion of potential non-financial costs is beyond the scope of this article.

5 Masset *et al.* in this *IDS Bulletin* explore the trends in the evaluation of projects that have outcomes that cannot easily be monetised or reported by a common measure.

6 Values are based on standard assumptions of longevity of usage of capital goods.

7 Complete explanations as to how the DiD results were derived are offered in Barnett *et al*. (2018).

8 Values calculated by the authors and Fink *et al.* (2016) for returns to education nearly come to the same figure. The authors based their calculations on the Montenegro and Patrinos (2014) rate of return values for Ghana and wage rate found in the household survey.

9 Costs are presented at the sectoral level since the underlying cost data were not traceable to specific outputs.

10 Using the rate of return reported in Banerjee *et al*. (2015) as 133 per cent for Ghana in an income-generating project through asset formation.

11 Multiple sources were used to obtain unit costs extending a child remaining in school for an extra year, including Baird, McIntosh and Özler (2011), Evans and Popova (2014) and McEwan (2012).

12 Various sources were used: Doherty and Govender (2004) for contact with health worker; White *et al*. (2011) for malaria; PATH and WHO (2016) and author calculation for vaccinations; Levine *et al*. (2006) for contraception.

13 Calculated as averages by the authors from a payment schedule.

References

Acharya, A.; Masset, E. and Saha, A. (2017) *Midterm-Preliminary Cost-Effectiveness Analysis*, Hove: Itad

Baird, S.; McIntosh, C. and Özler, B. (2011) 'Cash or Condition? Evidence from a Cash Transfer Experiment', *Quarterly Journal of Economics* 126.4: 1709–53

Banerjee, A.; Duflo, E. and Qian, N. (2012) *On the Road: Access to Transportation Infrastructure and Economic Growth in China*, NBER

Working Paper 17897, Cambridge MA: National Bureau of Economic Research

Banerjee, A. *et al.* (2015) 'Development Economics. A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries', *Science* 348.6236

Barnett, C. *et al.* (2018) *Impact Evaluation of the SADA Millennium Villages Project in Northern Ghana: Endline Summary Report*, Hove: Itad, http://itad.com/knowledge-and-resources/mveval/ (accessed 25 September 2018)

Clemens, M. (2012) 'New Documents Reveal the Cost of "Ending Poverty" in a Millennium Village: At Least $12,000 Per Household', *Center for Global Development*, www.cgdev.org/blog/new-documents-reveal-cost-%E2%80%9Cending-poverty%E2%80%9D-millennium-village-least-12000-household (accessed 13 September 2018)

Doherty, J. and Govender, R. (2004) *The Cost-Effectiveness of Primary Care Services in Developing Countries: A Review of the International Literature*, Disease Control Priorities Project, Working Paper 37

Duflo, E. (2001) 'Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment', *American Economic Review* 91.4: 795–813

Evans, D. and Popova, A. (2014) *Cost-Effectiveness Measurement in Development: Accounting for Local Costs and Noisy Impacts*, Working Paper 7027, Washington DC: World Bank

Fink, G. *et al.* (2016) 'Schooling and Wage Income Losses Due to Early-Childhood Growth Faltering in Developing Countries: National, Regional, and Global Estimates', *American Journal of Clinical Nutrition* 104.1: 104–12

Gallup, J.L. and Sachs, J.D. (2001) 'The Economic Burden of Malaria', in J.G. Breman, A. Egan and G.T. Keusch (eds), *The Intolerable Burden of Malaria: A New Look at the Numbers*, Supplement to Volume 64.1 of the *American Journal of Tropical Medicine and Hygiene*, www.ncbi.nlm.nih.gov/books/NBK2624/ (accessed 13 September 2018)

de Groot, R. (2016) 'Ghana LEAP 1000 Impact Evaluation: Overview of Study Design', *Innocenti Research Brief* 2016-01, Florence: UNICEF Office of Research, www.unicef-irc.org/publications/pdf/RBMethods_2016_01.pdf (accessed 13 September 2018)

Hanson, K. (2015) *What Can Education Systems Research Learn from Health Systems Research?* RISE Working Paper 15/003, Research on Improving Systems of Education, www.riseprogramme.org/publications/rise-working-paper-15003-what-can-education-systems-research-learn-health-systems (accessed 13 September 2018)

Levine, R. *et al.* (2006) 'Contraception', in D.T. Jamison *et al.* (eds), *Disease Control Priorities in Developing Countries*, 2nd ed., Washington DC: The International Bank for Reconstruction and Development/The World Bank Group

McEwan, P.J. (2012) 'Cost-Effectiveness Analysis of Education and Health Interventions in Developing Countries', *Journal of Development Effectiveness* 4.2: 189–213

Mitchell, S. *et al.* (2018) 'The Millennium Villages Project: A Retrospective, Observational, Endline Evaluation', *The Lancet* 6.5: PE500-E513, www.thelancet.com/journals/langlo/article/PIIS2214-109X%2818%2930065-2/fulltext (accessed 13 September 2018)

Montenegro, C.E. and Patrinos, H.A. (2014) *Comparable Estimates of Returns to Schooling Around the World (English)*, Policy Research Working Paper 7020, Washington DC: World Bank Group

PATH and WHO (2016) *Delivering Vaccines: A Cost Comparison of In-Country Vaccine Transport Container Options*, Seattle WA: Program for Appropriate Technology in Health and World Health Organization, www.path.org/publications/files/TS_opt_in_country_transport_rpt.pdf (accessed 13 September 2018)

Romer, D. (2006) *Advanced Macroeconomics*, Boston MA: McGraw-Hill

Sen, A.K. (1985) *Commodities and Capabilities*, Oxford: Oxford University Press

Tan-Torres Edejer, T. *et al.* (eds) (2003) *WHO Guide to Cost-Effectiveness Analysis*, Geneva: World Health Organization

Tollefson, J. (2015) 'Millennium Villages Project Launches Retrospective Analysis', *Nature*, 12 August, www.nature.com/news/millennium-villages-project-launches-retrospective-analysis-1.18157#/b1 (accessed 13 September 2018)

UNDP (1990) *Human Development Report*, New York NY: United Nations Development Programme

White, M.T.; Conteh, L.; Cibulskis, R. and Ghani, A.C. (2011) 'Costs and Cost-Effectiveness of Malaria Control Interventions: A Systematic Review', *Malaria Journal* 10: 337

World Bank (2018) *General Government Final Consumption Expenditure (% of GDP)*, https://data.worldbank.org/indicator/NE.CON.GOVT.ZS (accessed 13 September 2018)´

Abductive Reasoning to Explain Integrated Development: Lessons from the Multi-Method Evaluation of the Millennium Villages Project*

Dee Jupp[1] and Chris Barnett[2]

Abstract The use of mixed methods in impact evaluation often focuses on the triangulation of findings or using qualitative data to explain statistical assessments of impact (net effect). In this article, we argue for the benefits of combining and contesting *ex ante* etic-informed theory with *ex post* emic-informed grounded theory. Using the case of the Millennium Villages Project (MVP) evaluation in northern Ghana, we highlight examples of where emergent theory led to a deeper understanding that valued how local people view change. In the context of an independent evaluation that set out to assess the MVP model – and with little scope for the evaluation to influence or adapt implementation – the theories of change created abductively helped fill a theoretical void. We argue that this is particularly applicable for integrated programmes where *ex ante* theory may struggle to capture the complex interaction of different sector-based activities, overlook unintended effects, and undervalue local perceptions of change.

Keywords: integrated development, Millennium Villages, integrated rural development, multisector, theory of change, impact evaluation, complexity, local perspectives, participatory approaches, participant observation.

1 Introduction

Whilst many acknowledge the virtues of mixed methods research as 'the best of both worlds', the focus for mixing in international development evaluations[3] has often been on the triangulation of findings through the careful integration, complementation, and sequencing of methods.[4] This article describes the insights gained – within the confines of an external evaluation – by comparing *ex ante* theory with *ex post* theory and the combination of inductive and deductive modes of reasoning. This leads to the use of abductive (pragmatist, explanation-driven) reasoning to explain impact. The

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.162
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'**; the Introduction is also recommended reading.

article uses a real-life example of a longitudinal evaluation of the Millennium Villages Project (MVP) in northern Ghana and highlights the challenge of developing specific theories for integrated rural development projects – a point also highlighted by Masset (this *IDS Bulletin*). The MVP evaluation team combined theory-based impact evaluation (TBIE)[5] (such as that advocated by White 2009; Jimenez and Puri 2017) with naturalistic, contextual studies which generated grounded theory. We argue that combining *ex ante* theory-based approaches with *ex post* grounded (emergent) theory acknowledges the value of both, goes beyond simple triangulation of findings, and generates more complex-aware explanations of change processes on which to base future development interventions. The article also highlights the essentiality for evaluators to understand, value, and embrace different ways of generating knowledge and being willing to do so – no easy task for multidisciplinary teams.

2 Background to the independent evaluation

The impact evaluation of the MVP in northern Ghana was commissioned by the UK Department for International Development (DFID) as a multi-year external evaluation of the Millennium Village (MV) model. For DFID, an independent external evaluation was a condition of providing funding for the project, which started in 2012 and ran for five years. At that time, there had been little robust evidence to support the claims of the MV model as a means of achieving the Millennium Development Goals (MDGs) at scale, with much resultant controversy in social media and the press,[6] as well as between leading academics and policymakers (e.g. Clemens and Demombynes 2013).

As such, the evaluation was required to be independent, and operationally separate, in order to permit the MV model to be implemented as intended by the project team and using the project's intervention logic as the basis to verify results. This is in contrast to many complex-aware evaluations currently carried out that emphasise the continuous development and adaptation of theories of change in response to learning and feedback (Patton 2011; Andrews *et al.* 2015). Rather, the evaluation of the MVP was a TBIE along the lines advocated by 3ie and others, which has become a dominant mode of evaluation to measure and attribute the impact of interventions (White 2009, 2011). Such evaluations depend on the development of a robust theory of change which articulates the sequence of results of an intervention, including how the causal chain is expected to work and the contextual (external) factors that help determine change. Good theories of change are constructed through careful processes of incorporating stakeholder views and existing evidence, and benefit from the contributions of both quantitative and qualitative exploratory research. TBIEs may draw on different approaches such as theory-based evaluation (Weiss 1997), theory-driven evaluation (Chen 2012), or contribution analysis (Mayne 2001), but all subscribe to the notion that the theory is primarily developed *ex ante* and the evaluation is intended to verify the extent to which the theory matches what is observed.

By contrast, we present the case where the *ex ante* theory of change is critically compared with *ex post* grounded theory in order to enrich analysis of the change process. Drawing on the different traditions of research practice – and fusing inductive and deductive approaches – helps evaluators from different disciplines to interrogate findings and provide better explanations of change. This we explore in terms of abductive theories of change that examine outcomes, and seek to find the best possible explanation for these outcomes whilst retaining the proviso that the contributions and attributions *may* be true. As Reichertz explains:

> It is precisely in this quality of being a 'means-of-inferencing' that we find the secret charm of abduction. On the one hand it is a logical inference (and thereby reasonable and scientific), and on the other hand it extends into the realm of profound insight (and therefore generates new knowledge) (2014: 300).

We argue that, as such, the process of abduction is particularly important for integrated programmes where theory-driven approaches alone may struggle to capture the complexity of how different sector-based interventions interact, overlook unintended or unforeseen effects, and undervalue how change is perceived by the intended beneficiaries of the intervention.

3 The challenge of 'theoretical eclecticism'

The initial challenge for the evaluation team was to understand the theory of the MV model in order to design the TBIE. The *Terms of Reference* (ToR) for the UK DFID-commissioned evaluation of the MVP noted:

> The MV model provides an integrated package of interventions to lift a rural community out of poverty. Its central hypothesis is that a local 'big push' addressing the most immediate capital deficiencies in communities and households is a necessary condition for reaching a threshold required to move towards local resilience and self-sustaining economic growth.[7]

The 'big push' was designed to break the poverty trap[8] and this was the theoretical basis provided by the MV pioneers (UN Millennium Project 2005: 39). It is, in essence, the idea that massive scaling-up of resources to the poor will lead to 'take-off' which in turn will lead to a virtuous circle of self-sustaining growth. Many observers found this to be an insufficient theoretical basis and offered more robust theories, including that the MV model followed a 'market-oriented approach' (Broad and Cavanagh 2006), Sen's capability approach (Diepeveen 2008), and modernisation theory (Carr 2008). The variety of possibilities led Carr to describe this as 'theoretical eclecticism' and he wrote:

> Thus, the MVP, in the absence of overt theoretical considerations, plucks 'useful' aspects of theory from their contexts, and amalgamates them into a single, hybrid approach to development that contains points of access and purchase for academics and policy makers of all theoretical persuasions: in other words, the MVP has something for everyone (2008: 338).

Whilst links can be made to aspects of all these theories, the reality was that the MVP had itself articulated no clear overarching mid-range theory,[9] except loosely something based on the grand theory of the 'poverty trap'. In practice, the MV model was a collection of interventions designed to demonstrate how the MDGs could be achieved at a local level. Carr (2008) noted that 'Sachs and McArthur (2005: 347) stated clearly that the MVP was built upon the "core truth" that there were "known packages of effective and generally low-cost interventions" that can and should be applied to the challenges of extreme poverty'.

4 The hunt for theory

The external evaluation team searched for other 'theory' on which to base the evaluation, recognising that these had to be promulgated and owned by the project – not purely an external interpretation. Apart from the grand theory of the 'poverty trap' mentioned above, a couple of others emerged. First, a set of generic 'intervention logics' provided by the project showed the anticipated interconnections from inputs-to-outputs-to-outcomes and then to MDG impacts. These were not, however, specific to the northern Ghana MVP, made no reference to planned interventions and connections that would really occur, and were largely silent on the sequencing and synergies between different sectoral activities. Second, there was a logical framework (logframe) required of the project funders, DFID. This attempted to fit the complexity of the MVP into a simple, linear hierarchy from activities through to impact. While this was specific to the northern Ghana MVP, in essence it listed the intervention packages and MDG indicators by which their achievements would be assessed. This pared-down representation of the project did not do justice to its integrated nature and intended synergy – with no real detail on the assumptions and multidirectional interconnections. In summary, none of these options provided a basis against which to evaluate the MVP.[10]

The evaluation team decided to take a pragmatist view of the evaluation, combining a suite of evaluation methods to best answer the broad question of whether the integrated approach worked (or otherwise).[11] The decision to take a pragmatist approach did not diminish the challenge posed by the fuzziness around the programme's theoretical basis – one that reoccurred in dialogue with the evaluation's peer review group (PRG) throughout the evaluation's five years. Clearly a theory-based impact evaluation would require inferring a theory. But what would be the best way? And how to ensure that this remained true to the intentions of the MVP?

5 Combining deductive and inductive approaches

The most common way in which theories of change are developed are essentially deductively. This is where the team started, with *ex ante* theories of change developed through a deductive process informed by information from the project, and implicitly the team's mental models. By midline, for example, it was possible to create a series of

causal chains from a more detailed and nuanced examination of the data, as well as from the MVP's reporting. Plus, prior to the endline in 2016, another set of causal chains was developed in a workshop, drawing on actual interventions identified from the MVP progress reports, with some verification in the field, including with project staff. The limitations of this approach were nonetheless acknowledged by the team, noting that the theories of change have been 'split sectorally for clarity and ease of interpretation' and 'this approach has limitations and does not fully take into account the complexity of an integrated programme such as MVP and the added value [synergy] of multi-sectoral implementation' (Nelson *et al*. 2017: 3).

It was in the latter part of the evaluation that the team deliberately pitted the *ex ante* and *ex post* theories against each other. The Reality Check Approach (RCA)[12] – unlike all the other quantitative and qualitative streams within the evaluation – was not framed by *ex ante* theory, and used themes that emerged from conversations, observations, and experience during immersion in villages to construct grounded theory (in this case, *ex post*). This was an emic view of change experienced and perceived by people living in the MV communities themselves. As evaluators, it was important to undertake this comparison. This was not only because there had been a lack of clarity on the theoretical underpinnings of the project, but also because of the constraints imposed by being independent from the project and providing an accountability function with limited influence.

To illustrate the advantages of bringing deductive theory development and inductive theory development together, we discuss just two examples from the MVP evaluation: agricultural production, and income/consumption dynamics. The explanations for change developed for each of these phenomena benefited from exploring both *ex ante* (deductive) theoretical perspectives and *ex post* (grounded) theory perspectives and resulted in a more balanced interpretation of results.

6 Example 1: agricultural production

Before the start of the project, the MVP had identified maize production as a potential profitable crop for this area of northern Ghana. The project directly supported this by provision of subsidised improved seeds, a fertiliser advance scheme, tractor hire, and training on agronomic practice through a lead-farmer approach, which worked through farmer organisations (enabled by the provision of motorbikes and fuel, provision of additional agricultural extension posts, and salary 'top-ups' for these and the original extension officers) and indirectly through improved roads and telecommunications. The *ex ante* theory of change developed from the project intentions indicated that this (along with a focus on a few other selected potential crops) would lead to increased income, food security, and nutrition.

Based on the statistical (difference-in-difference) analysis of MVs and control villages (CVs), the evaluation finds that:

> [Overall,] the MVP has had a sizeable impact on agricultural production… with an approximately 38% rise over the period of implementation. This can be explained mostly by input increases (fertilisers, seeds, land, tractor rents and other animals/machinery for hire), with 74% of the productivity increase explained in this way… The land area dedicated to maize and beans[13] has increased considerably in the MV areas (Barnett *et al.* 2018: 166).

and

> the MVP had a large impact on food security, with a reduction in the number of households reporting not having enough food in the previous 12 months (*ibid.*: 126).

This seemed to validate the *ex ante* theory of change with the project inputs and activities being directly and wholly responsible for the improvements. But, was this the complete picture of what happened? The qualitative studies suggested a slightly different scenario. In 2013, there had been a major problem with the provision of fertiliser in time for maize production, tensions over repayment arrangements, and eventually an abandonment of this form of support. There was some evidence that tractors, seeds, and training may have been preferentially provided to larger landowners. But nevertheless, there was a palpable 'buzz' amongst less well-off farmers included in participatory rural appraisal (PRA) focus groups and those met during the RCA that people were getting good prices for maize (market price increased between 2012 and 2015)[14] and prices for the traditional crop of millet were decreasing. Previously, farmers in the area had grown millet to make their staple dish of *tuo zaafi* (TZ) and to brew alcohol. It was becoming common knowledge that maize could replace millet for both traditional purposes (TZ and alcohol). Millet has a longer growing season and is not harvested until late October/November and farmers soon heard that maize could be cultivated in a matter of three months (late April to early August), thereby freeing up the land for a dry season crop. The crop of choice was cowpeas, as it is sown soon after the maize is harvested (in July/August) and matures in about two months. Cowpeas are relatively drought-tolerant and the harvested dried beans are easy to store. By rotating with maize, farmers also benefit from the nitrogen fixation attributes of this legume.

Interestingly, the MVP had not articulated this potential in their bundle of intervention strategies and had not promoted cowpea production. As a consequence, the evaluation team did also not include this in their articulation of causal chains. The understanding of this phenomenon grew over time. By 2015, the midline RCA study noted that there was a clear distinction emerging between families who were able to grow cowpeas and those that did not. By 2017, the RCA study noted 'a slight improvement in the availability of cooked food each day in homes and more willingness to serve beans with "*tuo zaafi*" than previous years' (Jupp 2018: 48). Also, the PRA focus groups:

Box 1 The extraordinary rise of cowpea cultivation

Intrigued by so many people telling us how key cowpea cultivation had become in their lives and how many described a change in their outlook since they started to grow it, we tried to establish the drivers of this phenomenal uptake. Five years ago, cowpea was grown in riverine areas (e.g. Builsa North) but generally for consumption only as yields were not particularly good (less than three bags per acre) and no insecticides or weedkillers were used. People told us they began to hear about a private farmer in Yagba who was 'getting 15 bags from an acre' and wanted to find out how he did this and found that he achieved these yields by using chemicals bought from a single supplier in Bolga. Dealers quickly saw that demand for these chemicals was increasing and opened a number of outlets, permanent and at weekly markets, especially in Fumbisi. People copied the practice during the 2012–14 seasons and were excited by the yields they got and the fact that it filled an otherwise relatively unproductive farming period (October–December). Soon buyers were sending trucks from Kumasi, Tamale, and Techiman, and farmers shared with us that there is a 'big demand down south'. Large landowners also became interested in cultivating cowpeas, providing sharecropping opportunities too. People described the market as commercial and competitive and shared the advantages of storage if they can afford it.

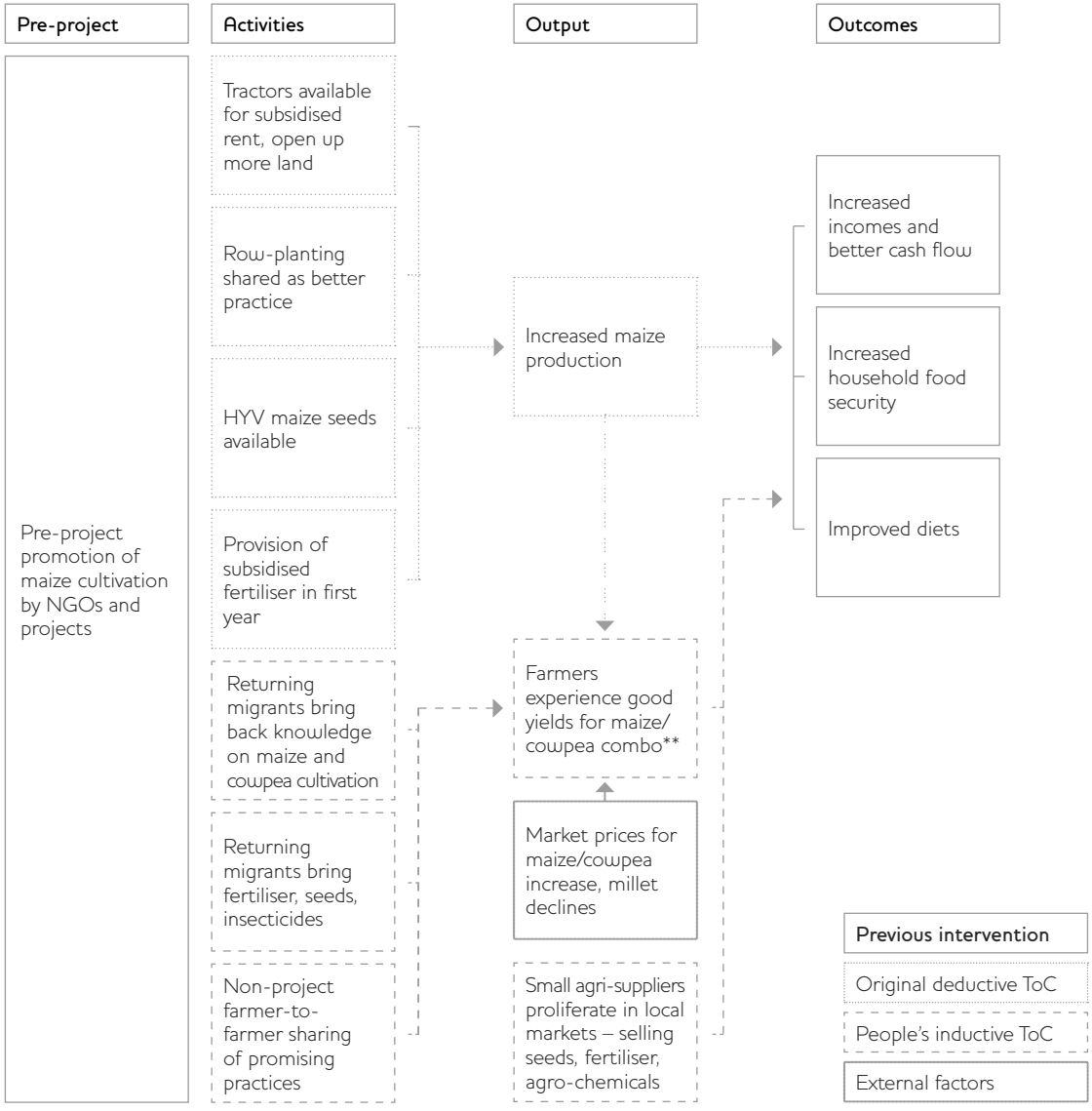
Source: Barnett *et al.* (2018: 119–20).

> indicated that MV farmers transferred skills they were taught about soybean cultivation to farming the higher-return cowpea crop. In contrast, farmers in those CVs [control villages] where cowpea has become a key crop often said they had picked up the requisite knowhow during their migration stints to the fertile so-called 'overseas' areas and along the White Volta, where they often hired themselves out as farm labourers on cowpea farms (*ibid.*: 99).

The RCA team were interested by the surprising rapid introduction and success with cowpeas over the five years that they had intermittently stayed with families and sought to understand how this had occurred. Box 1 describes what farmers shared.

The PRA added to this understanding by establishing that families were providing improved diets for their children mostly because they were adding cowpeas and eggs. Soybean takes longer than cowpeas to grow, takes longer to cook, and the taste is liked less than that of soybean, so the (limited) attempts to encourage its production by MVP were largely unsuccessful. Cowpeas are often now referred to as the 'hungry season crop'.

Figure 1 **Simplified schematic illustrating the development of the abductive theory**



Notes NGO – non-governmental organisation; HYV – high-yielding variety; ToC – theory of change.
**Maize/cowpea – two quick-growing, relatively drought-resistant crops that reduce risk from climate change, and increase productivity.
Source Authors' own.

Combining these insights and constructing a grounded theory of change alongside the *ex ante* theories of change, the evaluation team was able to abductively create a new theory of change (illustrated in Figure 1). The evaluation team concluded that the abductive best fit explanation was that the introduction of maize to substitute for millet was the key driver of change (for which the MVP can take some credit), but importantly it was the subsequent take-up of a second crop of

cowpeas in the dry season (due to factors beyond the project) that made an important contribution to income and food security.

7 Example 2: income and consumption dynamics

Our next example of *ex ante* and *ex post theories* combining to provide alternative explanations of phenomena concerns income and consumption data. Based on an analysis of the household survey, the midline evaluation report explained:

> We calculated monetary poverty rates using the methodology and the poverty lines used by the Ghanaian Statistical Service. Monetary measures of poverty are based on consumption-expenditure data at household level which include purchases as well as own-produced items and in-kind gifts from other households. Consumption is an overall measure of well-being as it includes the consumption of food, education items, health, non-food items and an evaluation of the value of the use of durable goods and housing. Households that fail to consume a minimum basket of essential goods are classified as poor (Masset *et al*. 2016: 45).

It went on to suggest that the project had had no impact on overall poverty but did have a significant effect on food poverty.

The qualitative strands (PRA and RCA) which complemented the survey showed that households employed more diverse, locally relevant descriptors of poverty/wellbeing in analysing changes in their conditions. Having cash in what had hitherto been a largely cashless society was highlighted in both PRA and RCA as an important element of 'not feeling poor', as was having family members who could migrate south for cash-earning opportunities in the dry season (*ibid*.: 48). The RCA study indicated that indebtedness was of grave concern amongst some poorer families with repayment difficulties leading to concomitant stress, concern for the future, and exclusion from further formal and informal loans (*ibid*.: 49). The PRA study facilitated people's self-generated descriptors of wellbeing from focus groups which included the multidimensional aspects of dignity, food security, personal health, educational attainment, tangible assets, and the ability to recover after financial shocks.

The endline report indicated that the project had had some effect on income, but expenditure remained unchanged. The endline report noted that:

> the fact that incomes increase while expenditure remains unchanged is counterintuitive and deserves some additional explanation. We would expect an extremely deprived population to increase their consumption as their income increases, but this is not what is observed from the survey data. One hypothesis is that the households in our study appear to have perceived the increase in incomes brought about by the intervention as temporary and have not adjusted their consumption upward (Barnett *et al*. 2018: 69).

Box 2 Funerals: a significant but underestimated expenditure?

Despite the MVP reportedly having made efforts to educate citizens on the impoverishing impact of funerals… and the need to control such expenditures, there has been little obvious impact on this deeply entrenched behaviour. As highlighted in earlier reports, people shared that surplus cash is often earmarked for funerals, which may be delayed for years if there are insufficient resources to '*honour the dead properly*'.… Arguably though, funerals can be viewed as an investment both to ensure the dead join the spirits of the ancestors and concomitantly can provide guidance and protection of the living as typified by this quote (from a PRA focus group discussion): '*[I]f these funerals are not performed, households will not progress and their ancestors will [cease to] help them in their farming and other endeavours*'. Both the RCA and PRA studies find that people say that funerals take up significant shares of surpluses in most households. During the midline RCA, people helped to construct the average cost for the two parts of the funeral ceremony (*Kumka* and *Juka*) and calculated a figure of around GHS1,200.

The PRA study found that especially in the Bul'k (Builsa) project area, funerals threaten to reverse much of the gains from improved agricultural productivity. The overwhelming majority of those interviewed felt that funeral expenditures are still rising. Large quantities of millet and maize are used to feed the throngs of mourners who congregate from near and far, but also in brewing the alcoholic beverages served for refreshment. Livestock expenditures are equally high, with households yielding to longstanding cultural prescriptions obliging them to make animal sacrifices to their ancestors. Furthermore, both the PRA and RCA studies find that funeral spending has increased as contemporary expectations have increased. The RCA and PRA researchers experienced over the years a trend to play modern music on (hired) hi-fi systems and to serve guests with bottled beverages rather than the traditional *pito* (a local beer brewed from millet and served lukewarm). A ban by the Bul'k paramountcy on serving (cheap) local gin at funerals (because many people end up intoxicated) has contributed to the shift to (more expensive) bottled beverages. Increasingly, the actual food served also comes in takeaway packs, which is considered a mark of sophistication. A participant at MV4 [a village under the MVP] noted [that] '*we are now integrating the Ashanti culture into ours*.' Others expressed the opinion that the rich have set new standards which they feel '*obliged to meet, to avoid disgrace*'.

Source: Barnett *et al*. (2018: 69–70).

Initially, this seemed counter to the RCA and PRA findings that suggest that some people are consuming more (e.g. marriages are bigger with larger sound systems, funerals are bigger with more gifts for guests, etc.), there is more alcohol consumption, more eating out, and more expenditure on TVs, renting DVDs, more motorbikes, and fuel costs. Box 2 describes the increase in funeral expenditure over the period of the MVP.

The explanations emerging from the qualitative streams had support in the research literature. For example, a study with the Kasena people who live in northern Ghana reported:

> Funeral rites take place sometime after burial of the deceased, the time lapse depending on the family circumstances – consensus among the immediate kin and elders about when to hold the rites and more importantly, the availability of quantities of foodstuff necessary for the rites (Awedoba 2014).

They describe these as the most important social events and 'the occasion for conspicuous consumption' which can be postponed for years.

Looking at the evidence abductively left the team puzzled. People had been keen to share that cowpeas had increased their cash incomes and that they had bought assets – notably TVs and motorbikes, and made improvements to their houses – but were these small investments (often made through credit repayments) too small to record a difference between MVs and CVs? These were all first-order purchases for those with cash profits from cowpeas. Second-order expenditure typically included that set aside for delayed social obligations (funerals and marriages) which, given the increasing lavishness of such celebrations, required savings from more than one crop season.

Eventually, the team settled on another explanation based on distinguishing between two types of expenditure: firstly, the consumption of current goods, which is employed in official poverty statistics (and which is not increasing in this case); and secondly, the purchase of durable goods and extraordinary expenditure, which is not included in official poverty statistics. This second type of expenditure (either as extraordinary goods or savings) *is* increasing – as supported by evidence from both the statistical analysis and the qualitative studies. Therefore, while households in the study area are perceiving the increased incomes resulting from the MVP as temporary, they are also altering their consumption patterns in ways that are underestimated by not being included in the expenditure used in official poverty statistics (such as gifts at funerals, which can be sizeable in the local culture) (Barnett *et al.* 2018: 69–70).

8 Concluding remarks

Of course, combining the perspectives of evaluator or project staff with a 'bottom-up' perspective of theories of change is not new, but it is also something that is not routinely done. If it is, it is often done with an

intention to improve *ex ante* theories of change. As Valters (2014) points out, the very process of developing theories of change

> can also create an illusion of serious reflection by being a superficial process of critical thought, where people who engage with the theories (donors as well as implementers) do not actually reflect sufficiently on how power dynamics change in practice and how local people see change happen.

In this way, the drive for theory-based evaluations – where evidence is primarily collected against the theory – can limit the ability to see beyond mostly confirmatory explanations of how change occurs.

While the advantages of mixed methods are often extolled purely in terms of improved triangulation and complementarity in evaluation, the team recognised value in conflict and contestation from sharing findings from different methodological perspectives. Greene, Caracelli and Graham describe this process in terms of 'the discovery of paradox and contradiction, new perspectives of frameworks, the re-casting of questions or results from one method with questions and results from the other method' (1989: 259). This requires contesting different theoretical lenses to the data with the intention of producing more thoughtful analysis, and leading to abductive (best fit) explanations;[15] which in the end would be more useful to understanding the theory of change behind the impact of the MVP.

This can generate important new insights. For instance, populist journalist Nina Munk followed the progress of a group of Millennium Villages for several years, and in her 2013 book *The Idealist: Jeffrey Sachs and the Quest to End Poverty*, she said a basic flaw of the MVP is that it is developed by academics living far away from the subject areas and with a poor understanding of local cultures, who do things such as promoting growing maize amongst people who have not historically eaten it or building a short-lived livestock market when there was no local demand. This is not wholly true and underestimates the contribution of the project: in the case of the northern Ghana MVP, maize *did* provide a breakthrough but the way in which it did was different to what was expected – i.e. mostly because it provided a cultivation window for the easy-to-grow, profitable, highly nutritious, and much liked cowpea. This insight was only possible from looking at a people's perspective of change and inductively improving on the deductive *ex ante* theory of change.

The examples provided in this article are just a 'taster' of many others the evaluation team encountered through purposely combining *ex ante* deductive and *ex post* inductive theoretical perspectives. The resultant abductive explanations require that all evidence is considered together and provides a better basis for making conclusions about what should be replicated or scaled up. Most ToRs for evaluation typically require the elaboration of theory as a pre-requisite, but we argue that they should also include a requirement to look at the emerging data *ex post*

through the experience of people who were intended beneficiaries of the intervention. The differences can be profound. Schindler, Graef and König, for example, studied causal chain explanations for agricultural upgrading strategies (UPS) and concluded:

> We observed that farmers and scientists have considerably different views on the positive and negative impacts of proposed agricultural UPS. While scientists focus mostly on direct causal impact chains of the UPS, the farmers consider indirect linkages that take their complex livelihoods into account (2016: 42).

They go on to explain that,

> while scientists assessed very high positive impacts on yield because of fertilizer application, the farmers also considered in their evaluation the high risk of lack of rain and chemical fertilizer application, which would increase yield failure even further during drought years compared to not applying fertilizer (*ibid.*).

However, this example is confined to the co-production of *ex ante* theory of change. In situations where external evaluation must be independent, the review of both *ex ante* (project-led) theory and *ex post* (people-led) theory fills an important gap.

It is in this way that we can get closer to providing better explanations of what happened and make better predictions of what might happen if aspects of the programme are reproduced. This is a particular challenge for those evaluating models of complex, integrated projects, where the multiplicity of interconnections are difficult to fully know from the start. It is even more crucial for evaluations that primarily have an accountability function, and where the main motivation is to 'test' and better understand whether the model of integration works – rather than continuously adapting the model. As Carr (2008: 336) notes in reference to MVP sites: '"known packages of effective interventions" may not capture the complex linkages between sectoral issues that result in the local challenges identified as problems by villagers'. Applying abductive reasoning to theories of change is one way to better understand these linkages, as well as challenge the confirmatory bias of our own mental models.

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 Technical Advisor, Empatika and External Research Associate, School of Development Studies, University of East Anglia.

2 Honorary Associate at IDS and Director of Technical Excellence, Itad, Hove, UK.
3 While the differences between research and evaluation are debatable and contested, this article focuses more narrowly on the application of different methods to the field of international evaluation. We distinguish 'evaluation' from the 'research' more broadly in terms of evaluation's focus on: (i) a more limited set of questions (essentially around understanding change due to a project or intervention); and, (ii) its express purpose of informing stakeholders (funders, implementers, etc.) in operational and policy decisions. By 'international development evaluations', we focus on those that take place primarily in low- or middle-income countries, with a strong focus on addressing different aspects of poverty and inequality.
4 Such as the sequential and concurrent designs for impact evaluation described in Bamberger (2012: 9–12).
5 TBIE combines an experimental or quasi-experimental design (measuring the net effect of an intervention) with a theory of change (using other methods to explain the link between project activities and the net effect, or impact).
6 Such as in *The Economist* (www.economist.com/feast-and-famine/2012/05/14/millennium-bugs), and correspondence in *The Lancet* (www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)60848-4/fulltext).
7 *Terms of Reference*, 22 September 2011 (DFID 2012: 1).
8 Broadly, this is an economic theory that suggests that there is a self-reinforcing mechanism which forces people to remain poor, and this is so binding that it requires a significant injection of capital for people to escape it.
9 Mid-range theory is defined here as a theory of change with limited scope that is able to explain a specific set of phenomena, in contrast to grand theory that might explain a phenomenon at a more abstract societal level.
10 Discussion based on an unpublished internal paper to the Peer Review Group, 'Responding to the PRG Comment on: Informing the Theory of Change with the Baseline Findings', 11 December 2013.
11 The evaluation design comprised four main strands: (1) a difference-in-difference (DD) methodology used to estimate impact, based on the difference in the change over time in the average outcomes between the project and in the comparison groups; (2) an Institutional Assessment that captured institutional change, particularly at the district and community levels; (3) an adapted form of PRA that was used to obtain feedback and insights on preliminary statistical findings from the perspective of different wellbeing groups; and (4) a reality check approach (RCA) that used a condensed immersion approach to better understand the realities of households and how they viewed change.
12 Immersive research where a team of researchers live in the homes of people in the community and informally interact, observing and experiencing daily life without using a preconceived evaluation framework (www.reality-check-approach.com).

13 'Beans' refers to cowpeas throughout.

14 Tables 7.1–7.3, www.e-agriculture.gov.gh/index.php/2014-07-22-14-39-46/agric-facts-and-figures-2015.

15 The core evaluation team included a range of professionals who describe themselves as economists, evaluators, statisticians, social scientists, and anthropologists. In discussing emerging findings, internal workshops and virtual meetings often involved forms of contestation that spanned methodological, theoretical viewpoints and worldviews – but with the aim of generating new insight.

References

Andrews, M.; Pritchett, L.; Samji, S. and Woolcock, M. (2015) 'Building Capability by Delivering Results: Putting Problem-Driven Iterative Adaptation (PDIA) Principles into Practice', in A. Whaites, E. Gonzalez, S. Fyson and G. Teskey (eds), *A Governance Practitioner's Notebook: Alternative Ideas and Approaches*, Paris: Organisation for Economic Co-operation and Development (OECD)

Awedoba, A.K. (2014) 'Wealth, Consumption and Migration in a West African Society. New Lifestyles and New Obligations among the Kasena, Northern Ghana', *Anthropos: International Review of Anthropology and Linguistics* 109: 45–55

Bamberger, M. (2012) *Introduction to Mixed Methods in Impact Evaluation*, Impact Evaluation Note 3, Washington DC: InterAction and The Rockefeller Foundation

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

Broad, R. and Cavanagh, J. (2006) 'The Hijacking of the Development Debate: How Friedman and Sachs Got It Wrong', *World Policy Journal* 23.2: 21–30

Carr, E.R. (2008) 'The Millennium Village Project and African Development: Problems and Potentials', *Progress in Development Studies* 8.4: 333–44

Chen, H.T. (2012) 'Theory-Driven Evaluation: Conceptual Framework, Application and Advancement', in R. Strobl, O. Lobermeier and W. Heitmeyer (eds), *Evaluation von Programmen und Projekten für Eine Demokratische Kultur*, Wiesbaden: Springer VS

Clemens, M. 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

DFID (2012) *PO 5603 MV-Evaluation: Terms of Reference, Impact Evaluation of a New Millennium Village in Northern Ghana*, ITT Volume 3, London: Department for International Development

Diepeveen, S. (2008) 'Putting Empowerment into Practice: Evaluating the Potential for "Development as Freedom" in the Millennium Villages Project', *Undercurrent: The Canadian Undergraduate Journal of Development Studies* 5.5: 6–13

Greene, J.; Caracelli, V. and Graham, W. (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis* 11.3: 255–74

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.D. Tamondong (eds), *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*, Exeter: International Development Evaluation Association (IDEAS)

Jupp, D. (2018) *Endline Reality Check Approach for the Millennium Villages Evaluation*, Hove: Itad

Masset, E. *et al.* (2016) *Millennium Villages Evaluation: Midterm Summary Report*, commissioned by the UK Department for International Development, Hove: Itad

Mayne, J. (2001) 'Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly', *The Canadian Journal of Program Evaluation* 16.1: 1–24

Munk, N. (2013) *The Idealist: Jeffrey Sachs and the Quest to End Poverty*, New York NY: Doubleday

Nelson, K. *et al.* (2017) 'Millennium Villages Evaluation: Sectoral Causal Chains', revised May 2017, unpublished, Itad, Hove

Patton, M.Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York NY: Guilford Press

Reichertz, J. (2014) 'Induction, Deduction, Abduction', in U. Flick (ed.), *The SAGE Handbook of Qualitative Data Analysis*, London: SAGE Publications

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

Schindler, J.; Graef, F. and König, H.J. (2016) 'Participatory Impact Assessment: Bridging the Gap Between Scientists' Theory and Farmers' Practice, *Agricultural Systems* 148 (October): 38–43

UN Millennium Project (2005) *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals*, London: Earthscan Publications

Valters, C. (2014) *Six Key Findings on the Use of Theories of Change in International Development*, Justice and Security Research Programme blog, London School of Economics, http://blogs.lse.ac.uk/jsrp/2014/08/18/six-key-findings-on-the-use-of-theories-of-change-in-international-development/ (accessed 7 August 2018)

Weiss, C.H. (1997) 'Theory-Based Evaluation: Past, Present, and Future', *New Directions for Evaluation* 76: 41–55

White, H. (2011) 'Achieving High-Quality Impact Evaluation Design through Mixed Methods: The Case of Infrastructure', *Journal of Development Effectiveness* 3.1: 131–44, https://doi.org/10.1080/19439342.2010.547588 (accessed 7 August 2018)

White, H. (2009) *Theory-Based Impact Evaluation: Principles and Practice*, 3ie Working Paper 3, New Delhi: International Initiative for Impact Evaluation (3ie)

Can Immersion Research Add Value in Understanding Integrated Programme Interventions?*

Dee Jupp,[1] David Korboe[2] and Tony Dogbe[3]

Abstract In the external evaluation of development interventions, beneficiaries are often involved to define priorities, and provide feedback and evaluation. This generally uses 'invited spaces' such as facilitated community meetings, focus groups, mobile phone-enabled feedback, and social audits. Where external evaluation must be both independent and separate from the project's own learning and adaptation processes, this can pose challenges. This article asks whether informal immersion in beneficiaries' 'own space' can provide insights beyond 'invited spaces' to enhance our understanding of how people experience development interventions, particularly where these interventions are integrated and complex. The article describes the inclusion of one type of immersion research, the Reality Check Approach (RCA) within the suite of qualitative and quantitative methods used in the longitudinal (external) impact evaluation of the Ghana site under the Millennium Villages Project (MVP). The RCA in this evaluation provided a means to spend concentrated time in beneficiaries' 'own space' without a project (theory-based) evaluation lens.

Keywords: immersion research; Reality Check Approach; beneficiary 'own space'.

1 Introduction

Cooke and Kothari (2001) noted three types of participation 'tyranny': the imposition of external systems of participation which override existing ones; the tyranny of group dynamics which favour consensus and are subject to co-option; and the tyranny of the approach which has squeezed out other ways of involving people and understanding their situation. These criticisms were subsequently strongly refuted in the context of participation as a transformative and empowering process, citing a growing body of evidence of the use of a range of innovative participatory approaches (Hickey and Mohan 2004). Unless specifically designed to provide a third-party learning function, external evaluation[4] is often by its nature required to remain outside

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.163
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'**; the Introduction is also recommended reading.

of the project, in order to provide an independent assessment of both the impact and processes described by the project design and business case. Such separation constrains the use of genuinely empowering participatory processes in evaluation, as these might be expected to influence the process and outcomes of the project.

The conventional means to involve beneficiaries in such external evaluations are typically through 'invited spaces' (Gaventa 2006) such as focus groups, interviews, specially convened participatory fora (e.g. as used in a range of participatory approaches such as participatory rural appraisal (PRA), social audits or Most Significant Change (MSC) workshops) – which mostly use project evaluation frameworks as a basis to solicit assessment. Since external evaluation is constrained by the requirement of non-interference, there is then little but 'invited space' to employ to involve beneficiaries. Rarely do external evaluations draw on 'popular space' (defined as arenas in which people come together at their own instigation – for example, to protest or for self-help and solidarity) (Cornwall 2004). This is often because beneficiaries continue to lack the agency and voice required to occupy 'popular space' (Mahmud 2007).

Given these limitations, there is more risk that the criticisms raised by Cooke and Kothari (2001) may be valid in external evaluation situations than, say, in adaptive project design and implementation. In such circumstances, immersion research may provide a middle ground between people's participation in 'invited spaces' and 'popular space', by enabling insights to be gathered into how people experience change. Does this have the potential to provide a sort of 'passive participation space' where the costs of participation for people are minimised and where the informality of the arrangement enables engagement with those who often eschew (or are excluded from) more public types of engagement?

The Reality Check Approach (RCA)[5] is one form of immersive research which has many applications, but in this case was integrated within the mixed methods longitudinal and external impact evaluation of the MVP in northern Ghana. The RCA involves researchers living with people in their own homes for a few days and nights in order to 'hang out' (Geertz 1998) with them, their neighbours, and others they meet during the course of ordinary days and nights. This informal interaction in people's 'own space' is intended to reduce power inequalities between people and researchers (with researchers 'experiencing' day-to-day life) and provides opportunities to combine listening with observation and direct experience.

The RCA was included in the suite of mixed methods for the evaluation of the MVP as a means to get closer to understanding people's perspectives and experience of change throughout the five years of the project. In this sense it was exploratory, but it was anticipated in the design and sequencing of the immersions with other evaluation approaches that it would also be able to provide alternative and emic narratives around change and the drivers of change. It was considered that this approach might complement the use of PRA with convened groups.

Whilst the PRA sessions were well facilitated and participatory, they were nevertheless 'invited spaces' (albeit at village level), to which the same cohort of participants were invited at each phase of the evaluation, and participatory analysis was largely focused on project processes and interpreting emerging findings on effects. By contrast, the RCA was not framed within an evaluation or project context but rather one where researchers wanted to immerse in day-to-day life and gather insights from this immersion. Further, in this evaluation, all researchers but three team leaders did not know they were in any way associated with the project[6] and were purposely primed to immerse in communities with open minds, and to focus only on how people experienced change and how they attributed change. This provided something similar to a double-blind approach, in that not only were the field researchers independent of the project, but the people they lived with, and interacted with, also did not filter their views through a project evaluation lens.

Immersion can add value compared with many other qualitative approaches, as it provides opportunities for immediate triangulation – comparing what people say has changed with direct observation and the researchers' own experience, for example. Those who habitually undertake immersion studies also note during critical reflection post-immersion that engaging with people through shared experience and using the context to stimulate conversation leads to a perception of greater authenticity in the findings (and by inference, a reduction particularly in social desirability and habituation bias). Conversations are natural, two-way, and can be returned to as new insights emerge over the course of the immersion. Immersion offers opportunities to interact with those who may eschew 'invited spaces' or 'popular spaces', or who may not be able to participate in such spaces, thereby enabling a wider net of perspectives for triangulation. Living with people in their own space leads to opportunities to build trust which cannot be so easily built in 'invited spaces' with fixed meeting places, time constraints, and external agendas. It provides opportunities to explore unexpected insights and to iterate findings through being 'on the spot' over time,[7] including day and night, and on different days of the week. It enables people to participate without disruption to their daily activities and income-earning as researchers accompany and participate in chores too.

Immersion research nevertheless has limitations, including scale, selection bias, and researcher bias. In the case of the MVP Impact Evaluation, the RCA researchers lived with only 20 households, about three households in each of six villages (four 'treatment' Millennium Villages (MVs) and two comparison villages). However, households in northern Ghana are often quite large, usually multi-generational, and on each of the three occasions, the researchers lived with the households and were able to 'hang out' and have insightful interactions with an estimated 1,500 people comprising a wider demography including neighbours, local tradespeople, local leaders, and service providers such as school teachers and health staff. Given the size of the MVP, which

covered 34 villages and 30,000 people, the RCA only comprised a small number of households in context-specific locations.

The immersion was purposely with poorer households, although, as noted, there were important interactions with other socioeconomic groups and service providers. To mitigate some aspects of selection bias, researchers navigated the discussions with villagers and households on the proposed living arrangements themselves and avoided gatekeepers or intermediaries. Families were purposely located both in the centre and periphery of villages. Each immersion was also quite short (four nights and five days so over the course of the five years, 12 nights with each family).

Other limitations inherent in immersion research such as concerns about the positionality of researchers and differential power relationships between researcher and families, and within the family and community were discussed extensively during the training of researchers where means to mitigate such limitations were practised through simulations. Plus, like all ethnographic-based work, efforts were continuously made to critically review practice and exercise reflexivity. The RCA is always carried out in teams rather than as individuals (although immersion is individual) in order to reduce researcher bias, to widen the net of perspectives, and to ensure critical reflexive practice.

2 Insights gained from immersion in beneficiaries' 'own space'

Both the MVP's own monitoring system and that adopted by this third-party, external evaluation sought to measure achievement against pre-determined indicators, primarily the Millennium Development Goals (MDGs), but also those presumed to represent the results of the package of interventions. This meant that the emphasis was on measuring what was intended and was inherently a normative and, in this case, a purely externally constructed project view.

As mentioned above, the RCA interactions eschewed a project lens and looked at change more broadly. In the third and final immersion in 2017 (six months after the official closure of the MVP), the RCA researchers used the 'hanging out' methodology to purposely reflect with the families they had lived with over the previous five years as well as with the range of other study participants as described above, on change (good change and bad change) and the perceived drivers of those changes. As this was the third immersion, researchers were able to share the experience of many of these changes through using local transport and roads, accompanying children to school, accompanying patients to health facilities, eating with families, taking produce to market, experiencing change in agricultural practices, sleeping in their homes, using water and sanitation facilities, and so on.

The insights generated through informal conversations, shared experience, and observations were brought together immediately following the immersions during intensive sessions with each village

research team in turn, where details of the changes and their significance to people were compiled without filter and critical reflexivity was practised. These 'downloading sessions' were followed by all village research teams coming together for a sense-making workshop where findings were compared across locations and study participants. Immersion in the data set comprising records of conversations, observations, personal experience, photos, and artefacts led to a search for emergent themes which were coded and charted using conventional grounded theory approaches, with special attention paid to disaggregating themes by study participant categories (primarily gender, age, and socioeconomic categories[8]).

Interestingly, the three most significant changes identified across categories of adults were the same, although the reasons for the significance were often nuanced. Furthermore, people rarely noted any other change beyond these three. Based on multiple conversations and frequency of mention, the three key changes for adults of all ages and across socioeconomic categories were: (1) electricity connection, (2) cowpea cultivation, and (3) increased mobility and connectivity.

The MVP had prioritised specific 'proven' interventions, including in health (new and rehabilitated health centres and a range of preventative health-care interventions), schooling (new and rehabilitated schools, teachers, and resources), road improvements, and agriculture (introduction of maize, provision of inputs such as seeds and artificial fertiliser, and tractor hire) as an engine for economic growth. Significantly, people's views of what was significant change embraced social benefits as much as the economic benefits emphasised by the project. So, for example, the benefits of electricity connection were largely seen in terms of social benefits – household connections for home lighting, TV, and easing the burden of milling for home consumption, as well as ensuring that social services were resourced because teachers and nurses would not stay in villages without electricity. Cowpea cultivation was significant, not only because it provided additional cash (in contexts which increasingly require cash) in an otherwise unproductive season, but as a means to halt the seasonal migration for work which split families for many months every year. In terms of mobility and connectivity, it was the increase in what are locally referred to as *motorkings* (small three-wheeled motorised trucks used to transport goods and people) and private motorbikes that was noted and primarily viewed in terms of time-saving and convenience for recreational as much as economic purposes. The significance of the increased use of mobile phones was also couched mostly in terms of social connectivity, and only secondarily as an opportunity to access mobile money, and not in terms of accessing work or trade opportunities.

The following examines each of these three key changes in more detail, based on insights from the immersion in beneficiaries' 'own space', how these added to the external evaluation, and how they could have been useful for design of the intended further roll-out of the MVP, had this actually happened.

3 Differing priorities and views of integrated development

Under the MVP, electrification was framed in terms of an engine for economic development and more efficient service provision, and the project lobbied the Electricity Company of Ghana Ltd to prioritise the MVs in their National Electrification Scheme (NES) and in their rural electrification plan. Through conversations during the immersion, it was clear that connection to electricity was perceived as a major sign of development by adults of all ages, as well as children, with comments such as 'We are like towns now'. Families enjoyed light at night (mostly keeping the lights on overnight, saying that they liked to feel safe, to attend to children in the night, and that keeping lights on at night had become the village 'norm'), better milling facilities for personal use (not commercial use), and buying cold drinks from local shops.

But, without doubt, the most significant change referred to widely and observed in the final immersion was access to TV. The experience of walking round the villages at night in 2017 was very different from previous immersions in 2013 and 2015, when women and children went to bed soon after dusk, and men tended to gather in bars or small shelters to chat, but also retired to bed early. Many houses now had their own TV, usually bought second-hand with cash (around GHS150 (£24), satellite dish around GHS250 (£40)). People shared that it gave them something to enjoy, especially in the evenings, as illustrated by comments made by the elderly aunties in one home: 'We now play in the evenings and go to bed later'. Others noted that it made them feel 'connected' with the world outside their village.

The change also brought concerns. Many parents and school teachers worried about children watching TV late and oversleeping or dozing in class. Whilst some families liked neighbours coming around to watch programmes, others resented their loss of privacy. Bill-payers (both men and women) were increasingly concerned about the costs of their electricity, and many shared that they were fearing disconnection, something they thought they would feel acutely, having become accustomed to lights and TV. Whether they were conservative in their use of electricity or not, the costs of connection (GHS40–120), electricity consumption (minimum GHS2–15/month), and the need for bulbs (GHS2.5 each, which last about a month) were considered major additional cash expenses.

As well as the largely social benefits of electricity at household level, the teachers and nurses interacted with during the immersion explained that they would not have relocated to the villages if there had been no electricity, and where quarters had been provided without reliable electricity connections, these were not occupied. In their words, it meant that they could 'have fridges, satellite TV, and computers',[9] with some saying that this was essential as 'social life in the village is difficult'.[10] The immersion confirmed that they had very little interaction with villagers outside of school or work.

A number of new mills have been established or old diesel generator-powered mills have been converted, mostly through non-metered (illegal) connections. A project lens might conclude that this was demonstration of enhanced economic activity, but spending time at these mills 'hanging out' revealed that only small quantities are brought for milling for personal use just as before, and not for processing for on-selling by the customer. Women shared that 'electric mills grind flour softer and nicer'[11] and there is 'no smell of diesel'[12] compared to the past. Some shops had invested in fridges, but observation indicated that these are always filled exclusively with cold drinks.

In terms of raising agricultural productivity for economic gain, the MVP had decided *a priori* based on expert advice to concentrate on four promising crops; maize, rice, acacia, and mango. Extension programmes were designed to promote the cultivation of these crops including demonstration plots, subsidised seed and saplings, and new fixed-price tractor services. However, it was cowpea cultivation that was cited by men and women as the second most important change and the most important economically they had experienced, as illustrated by the quote from a young farmer in 2017: 'If you are not part of the cowpea business you are dead already'.[13] It was very evident from the 2017 immersion that the practice had grown phenomenally since 2013, with observation of stocks of cowpeas in people's homes and the addition of cowpeas to TZ[14] to at least one of the meals taken each day, when previously researchers and families had eaten TZ alone.

In conversations, men and women were clear that this has not been driven by external development programmes, but through farmer-to-farmer diffusion based on a single role model (whom different people in villages far from one another cited) who, in 2013 demonstrated hugely increased yields using insecticides and weed killer. An external view suggests the MVP's promotion of maize may have facilitated this change because maize has a shorter crop period compared with millet, the traditional crop commonly observed in immersions in 2013, allowing for a second crop (cowpeas) to be grown. However, people themselves felt that increased market availability of agricultural inputs were what enabled the change. Immersion in the villages indeed demonstrated the rapid rise in retail of agro-chemicals with new outlets along roadsides, market stalls in weekly markets which had not been there in 2013 and 2015, and a proliferation of discarded agro-chemical containers littering paths to fields surrounding the villages. The substitution of traditional millet growing with maize was also explained by people in terms of increasingly reduced millet yields, and families shared that it was also welcomed because millet was considered a labour-intensive crop, especially at maturation, when flocks of birds have to be kept away. The ease of access to agro-chemicals in local markets also meant that farmers could grow maize and cowpeas on their own without needing others to help them with clearing the soil and weeding, as they had done with other crops (requiring time-consuming, traditional reciprocal arrangements), or having to pay for labour. Those

in both project and comparison villages, who indicated that they felt better off than when researchers met them before in 2013 and 2015, attributed this almost entirely to cowpea cultivation, and many shared that they had been able to purchase assets such as tin roofing, extended their houses or, of course, bought TVs and motorbikes with the profits.

Economic gains were only part of the reason that people felt cowpeas were so significant. In some villages, cowpea cultivation was said to have helped reverse the traditional seasonal migration of both men and women that used to take place during the dry season. People shared, 'There is no need to go South now as the income from cowpeas is good and we have electricity'.[15] This was valued as a means to keep the family together and to avoid the stress created by concern for the safety of those migrating. Women shared that because they were key in harvesting and processing cowpeas, they were able to use small stocks as their own personal savings, and to sell small quantities for cash for items such as personal toiletries, which had not been possible before.

In addition to electricity and the cowpea cultivation discussed above, people also cited mobility and connectivity as a key change in the area. The MVP framed improvements in road infrastructure as part of its integrated approach, in terms of providing all-weather connectivity to make roads useable by haulage trucks to convey agricultural products out of the MV. Much like electricity provision, the MVP did not directly invest in this but lobbied the national and local government to prioritise the construction and maintenance of a number of key roads servicing MV areas.

Immersion in villages revealed a conundrum, however. Men and women alike highlighted increased mobility as a key change – like other changes viewed through a project lens, this could be attributed to the improved roads. But conversations based on shared experience between families and researchers often turned to the poor state of roads which had been rehabilitated with many collapsed culverts and dangerously deep potholes. One particular road was described as 'much worse than before'. Insights were gained by observation of the massive increase in *motorkings* and private motorbikes. The increase was observable in both project and comparison villages. The source of these bikes was a British–Ghana venture established in Tamale in 2007, with Chinese expertise to build and sell *motorkings* initially, and later motorbikes.

People explained, and researchers observed and experienced, that both types of vehicle can deftly navigate broken culverts and deep potholes and can access quite rudimentary paths. People also explained that the proliferation was enabled by the substantial increase in disposable cash (such as from profits from cowpeas) and easy credit terms provided by the local company. Men and women indicated that these vehicles were not expensive to run and provided a much more flexible and reliable option than the market trucks or pushbikes of the past, being both convenient and time-saving.

Living in the homes of families who had purchased motorbikes in the year leading up to the third and final immersion enabled observation and discussion around the use of motorbikes, augmented by observation and chats with others. Most journeys were for social reasons – to meet up with friends and relatives, recreational, and accessing fields – and less used for accessing markets for buying or selling.

The anticipated increase in haulage trucks envisaged by the MVP was not apparent, except on a few short stretches of the roads which were passable. On one major artery which had been rehabilitated through MVP efforts (as confirmed from signage seen by researchers), subsequent deterioration was so bad that market trucks could no longer pass, and the weekly market had halved in size as people preferred to access a different market using a different route. Women market-sellers from outside and families we lived with predicted the market would soon cease altogether.

In addition to physical mobility, people shared that they felt more connected due to mobile phones and this was almost entirely viewed in terms of social relationships. Tigo, the preferred service provider, was only in two MVP locations where RCA immersion was undertaken, and which initially had a partnership with MVP and installed masts in these two villages. In other villages, better signals were obtained from Vodafone and MTN. Economic benefits were only cited in areas with provision of good MTN signals which enabled access to mobile money.

4 Discussion

The MVP describes itself as bottom up, participatory, and community-led: 'Participatory, community-led decision-making is central to the way Millennium Villages work and is also fundamental to sustainability'.[16] As noted by Carr, however:

> Descriptions of the MVP as a 'bottom-up' approach are questionable, given the project's reliance on pre-conceived definitions of problems and pre-packaged solutions to address poverty at the village level. These pre-conceptions present serious challenges to understanding the actual problems faced by the people living in these villages (2008: 334).

The MVP did indeed make use of some 'invited spaces' to refine interventions. During the project, there was one example of 'popular space' use comprising a protest made by farmers over a failed scheme to provide fertilisers for maize production. These were provided too late, yields suffered, repayments could not be made, and farmers were harassed. Many farmers did not want to join the scheme the following year and it was dropped. Although the project convened community and group meetings, these were largely to provide information and organise interventions, and there was no apparent mechanism within the project to gather ongoing feedback or beneficiary assessment – nor obvious attempts to empower beneficiaries to actively participate and

influence the implementation of the project. Had the MVP adopted a more effective participatory approach to project management and adaptation, there may have been opportunities for the external evaluation team to have interacted with community groups and to learn about the progress and impact of the programmes through their lenses.

As the external evaluation used theory-based impact evaluation (White 2009) principles, this meant achievements were assessed primarily in terms of project objectives and its underlying Theory of Change. The evaluation methods used mostly involved 'invited spaces', albeit at village locations (such as household surveys, and facilitated group meetings). The inclusion of immersion research complemented this approach by not being constrained by the evaluation framework. Instead, it used a simple 'hanging out' methodology in beneficiaries' 'own space', together with grounded theory. In doing so, it could provide insights into what changes were important to people, such as highlighting how social meaning was mostly overlooked in project theory and design. It also prompted areas of enquiry and analysis which might have been neglected without the combination of conversations, observation, and experience inherent in and valued by 'hanging out'.

One important insight is that the three key changes highlighted by people above (electricity, cowpea cultivation, and mobility) were not directly funded by the MVP. From the outset, the MVP had intended to leverage electricity connections and improved roads as important enabling elements of the integrated approach. The MVP did not promote cowpea cultivation and did not noticeably react to the (concerning) rise in agro-chemical use. In terms of wellbeing and what development means, people also shared very different views than those perpetuated by the MVP; having a light on all night, owning a TV, a motorbike, and phones, and being able to consume cold drinks made important differences to the quality of people's lives. These also provided opportunities for social interaction – which all constitute significant contributions to a sense of wellbeing.

The experience of this immersion research suggests the desirability of including processes to connect with people's experience of the interventions made in their name on a regular basis. Had the immersion research been undertaken as part of project implementation, and not only as an explicitly independent external evaluation, the project could have benefited from being continuously fed these sorts of insights to improve the adaptation and sequencing of interventions within the integrated programme. Given the original intention that the external evaluation would determine whether further investment and up-scaling in the MVP should be considered, the immersion research could have also provided important insights into future programming by unpacking the normative input-driven model, and offering people's perspective unshackled from sponsor bias.

Whilst this is not new, the practice of engaging in beneficiaries' own spaces is often undervalued. Twersky, Buchanan and Threlfall (2013) note:

> In bypassing the beneficiary as a source of information and experience, we deprive ourselves of insights into how we might do better – insights that are uniquely grounded in the day-to-day experiences of the very people the programs are created for.

They suggest that unlike the private sector and their customers, in the development sector, it is easier to ignore the beneficiaries who often 'express gratitude for even subpar effort' (*ibid.*).

Integrated programmes such as the MVP are intended to address problems in a holistic manner and, by inference, acknowledge that people do not experience life through a series of sectoral lenses (such as separately for health or education). However, if such projects fail to consider people's experience and their value given to change, they merely co-ordinate different sector-based interventions with the 'hope' that there is some synergistic benefit; rather than recognise the importance of sequencing interventions, moving at the pace that people want, and adapting to new emerging possibilities and challenges. Each intervention within an integrated approach creates a series of ripple effects which need to be adjusted to. Immersion studies can help identify these effects early because of the opportunities provided, by combining listening with observation, and experience within beneficiaries' 'own space', with the intrinsic advantage over 'invited spaces'.

The use of immersion research described in this article may help address some of the criticisms of participatory practice levelled by Cooke and Kothari (2001) as noted in the introduction. The informality and 'hanging out' in beneficiaries' 'own space' has the potential to address many of the issues raised about co-option, unhelpful use of power, outside agendas, and problems encountered by interaction only in 'invited spaces'. Furthermore, it derives value from the researchers' shared experience and opportunities for observation, neither of which are emphasised in other qualitative methods. Conventional ethnography is often limited to a single researcher spending extended periods of theory-led immersion research in the community. There are benefits of having many researchers undertaking immersions concurrently, and without theory or evaluation framework lenses. Short immersion time and scope can arguably be trade-offs for insights which may be hard to gather from 'invited spaces'. Immersion research provides an informal means of engagement which addresses much of the concerns about the disincentives and costs to participate. Where empowering forms of participation are not undertaken because of poor participation practice, because of the required independence of external evaluation, or because people actively eschew public participation, immersion research may indeed provide a middle ground or 'passive participation space'.

There is a well-known African proverb, 'Only the one sitting on the anthill knows the ants are biting'. When researchers immerse in villages and live with families, they can at least get some taste of the ants biting and some insights on emic perspectives.

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 Technical Advisor, Empatika and External Research Associate, School of Development Studies, University of East Anglia.

2 Independent consultant.

3 Participatory Development Associates Ltd, Accra, Ghana.

4 External evaluation is defined for the purpose of this article as an independent function, reporting separately to the funder to provide rigorous, impartial evidence primarily for accountability purposes.

5 www.reality-check-approach.com.

6 Also reducing 'sponsor bias'.

7 While short compared to many ethnographic studies, these are longer periods of interaction than focus groups, interviews, and participatory approaches such as PRA.

8 Including categories derived through PRA wellbeing rankings.

9 Pers. comm., July 2017.

10 Pers. comm., July 2017.

11 Pers. comm., July 2017.

12 Pers. comm., July 2017.

13 Pers. comm., July 2017.

14 *Tuo zaafi*, a cooked very thick porridge of maize and water.

15 Pers. comm., July 2017.

16 http://milleniumvillages.org/the villages/.

References

Carr, E.R. (2008) 'The Millennium Village Project and African Development: Problems and Potentials', *Progress in Development Studies* 8.4: 333–44

Cooke, B. and Kothari, U. (eds) (2001) *Participation: The New Tyranny?*, London and New York NY: Zed Books

Cornwall, A. (2004) 'Introduction: New Democratic Spaces? The Politics and Dynamics of Institutionalised Participation', *IDS Bulletin* 35.2: 1–10, http://bulletin.ids.ac.uk/idsbo/article/view/1099 (accessed 14 September 2018)

Gaventa, J. (2006) 'Finding the Spaces for Change: A Power Analysis', *IDS Bulletin* 37.6: 23–33, http://bulletin.ids.ac.uk/idsbo/article/view/898 (accessed 14 September 2018)

Geertz, C. (1998) 'Deep Hanging Out', *The New York Review of Books* 45.16: 69–72

Hickey, S. and Mohan, G. (eds) (2004) *Participation: From Tyranny to Transformation? Exploring New Approaches to Participation in Development*, New York NY: Zed Books

Mahmud, S. (2007) 'Spaces for Participation in Health Systems in Rural Bangladesh: The Experience of Stakeholder Community Groups', in A. Cornwall and V.S. Coelho (eds), *Spaces for Change? The Politics of Citizen Participation in New Democratic Arenas*, London and New York NY: Zed Books

Twersky, F.; Buchanan, P. and Threlfall, V. (2013) 'Listening to Those Who Matter Most, the Beneficiaries', *Stanford Social Innovation Review*, https://ssir.org/articles/entry/listening_to_those_who_matter_most_the_beneficiaries (accessed 7 June 2018)

White, H. (2009) *Theory-Based Impact Evaluation: Principles and Practice*, 3ie Working Paper 3, New Delhi: International Initiative for Impact Evaluation (3ie)

This page is intentionally left blank

Learning About Integrated Development Using Longitudinal Mixed Methods Programme Evaluation*†

Emily Namey,[1] Lisa C. Laumann[2] and Annette N. Brown[3]

Abstract Globally, millions of children are living without parental care. Families experience multiple challenges that lead to family–child separation, with financial hardship a common theme. The integration of interventions to strengthen families socially, emotionally, and economically is therefore a priority, but requires knowing which combination(s) of interventions might work for which households. This article reflects on an ongoing evaluation of two projects in Uganda that implemented integrated family and economic strengthening interventions with families at risk of a child being separated and those where a separated child was reunited. We discuss how the practical realities of integrated programming influence research design options and our attempts to mitigate the potential limitations of a non-experimental design with the use of a mixed methods approach.

Keywords: evaluation, integrated development, child protection, economic strengthening, family strengthening, separated children, reintegration, mixed methods.

1 Background

1.1 Description of the problem

Millions of children around the world are living in situations without parental care (EveryChild 2009; Petrowski, Cappa and Gross 2017). These children might live in residential care; in alternative, family-based care; in child-only households; in juvenile detention; on the streets without their parents; or with their employers/exploiters. While in some cases parental care is not possible or in the child's best interests, children outside of parental care are more likely to be neglected and exposed to abuse and exploitation, are more vulnerable to health and behavioural problems, and face challenges in developing relationships, a sense of identity, and self-esteem (Bakker, Elings-Pels and Reis 2009; Berens and

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.164
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'**; the Introduction is also recommended reading.

Nelson 2015; Browne 2009; Csáky 2009; Dozier *et al.* 2012; Tobis 2000; Wedge, Krumholz and Jones 2013).

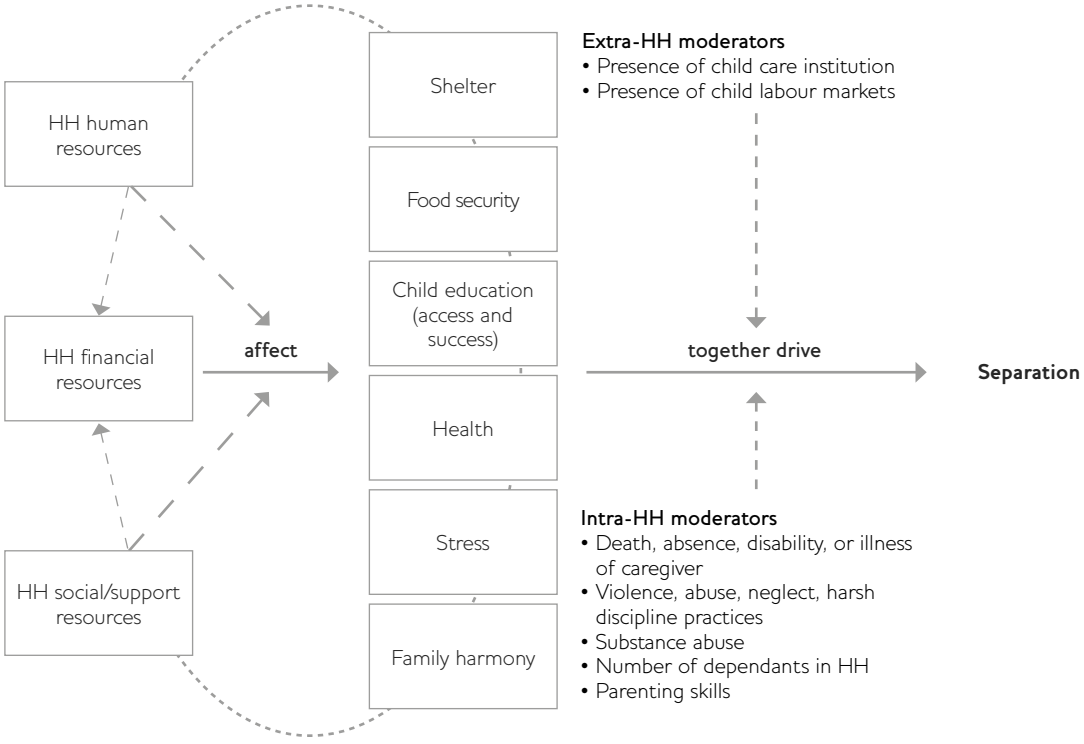
Assessments by practitioners suggest that families experience multiple challenges that can lead to the placement of a child in institutional care or domestic service, or lead a child to the streets (Delap 2013; EveryChild 2009; Laumann 2015). A lack of financial resources within the home to address children's needs is a common theme. Household economic interventions that aim to reduce poverty and build financial resources can therefore supplement traditional social work-driven efforts to help prevent family separation, support children's reintegration in family care, and improve outcomes for children (Chaffin and Kalyanpur 2014; United Nations General Assembly 2010). However, social work practitioners, with skills in psychosocial and emotional support, report that they lack knowledge and skill in addressing economic matters (Chaffin and Kalyanpur 2014), particularly in low-resource environments where government assistance programmes are limited or non-existent.

At the same time, practitioners in the economic/livelihoods domain with expertise in material support, asset creation, income-generation activities, and employability skills typically work at a broader level and often lack experience working with individual families on the verge of breakdown. Understanding how to successfully integrate protective family social and economic strengthening interventions is therefore a priority, as is bridging the gap between economic development and child protection practitioners (Boothby *et al.* 2012), particularly given the adoption of the Guidelines for the Alternative Care of Children by the United Nations General Assembly in 2009[4] which has propelled state and civil society action on separated children.

1.2 Introduction to the study

In 2014, under the FHI 360-managed[5] Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) umbrella, the United States Agency for International Development's (USAID) Displaced Children and Orphans Fund commissioned the Family Care[6] project. The goal of the Family Care project is to develop evidence and guidance for how economic strengthening interventions can help separated children return to and remain in their families and to help highly vulnerable families stay together. A principle source of this evidence is a study that includes the evaluation of two projects funded by Family Care that integrated family social strengthening and household economic strengthening. Ultimately, the Family Care team will combine the evaluation evidence with a literature review and other data sources to develop global guidance on the topic, including how to match families with the forms of economic strengthening activities most appropriate to their needs and circumstances, and how these activities can contribute to family strengthening efforts. In this article, we present the considerations that influenced our evaluation design, highlighting the design elements and approaches we incorporated to better assess the integrated, cross-sectoral nature of the projects.

Figure 1 Concept map of the common drivers of family–child separation



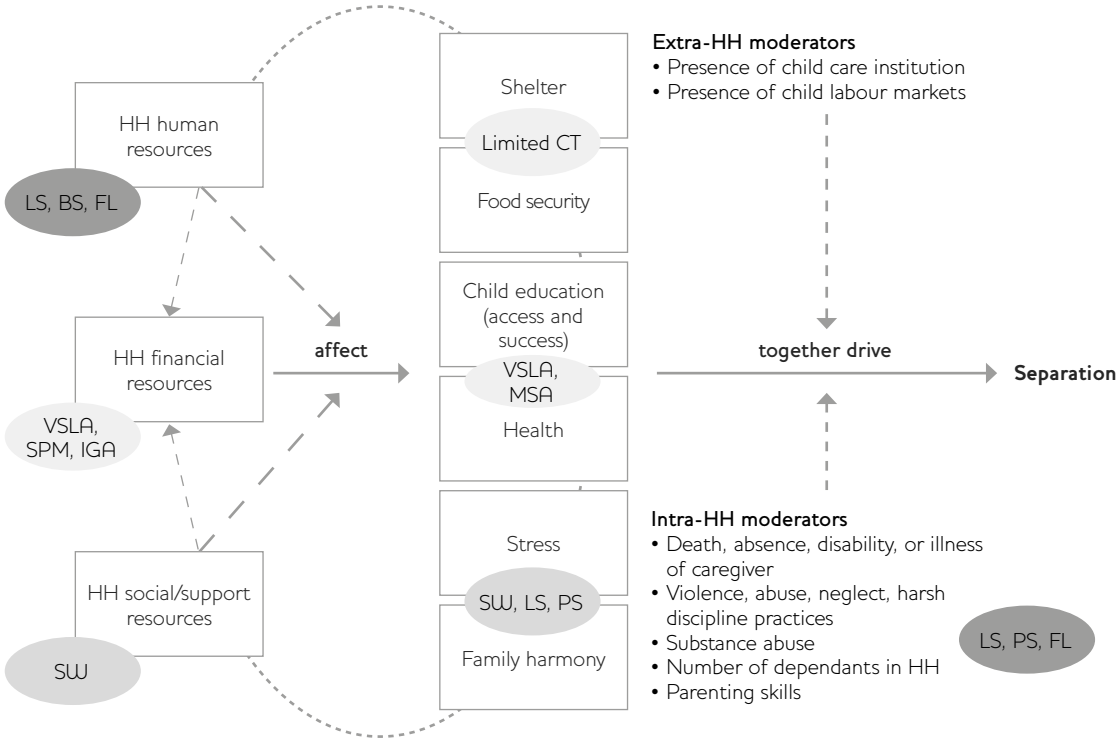
Note HH – household.
Source Authors' own.

1.3 Theory of change

Prior to developing an evaluation framework, and indeed, prior to issuing a solicitation for the projects we would evaluate, we needed to understand the drivers of family–child separation and how family social and economic strengthening activities might be expected to mitigate them. We began with the drivers identified in the literature (Delap 2013: 13; Laumann 2015) and validated and supplemented these with learning from field visits with practitioners working on deinstitutionalisation and reunification in Uganda and Rwanda. As illustrated in Figure 1, there are drivers related to both economic and familial characteristics of households – from inadequate shelter and food insecurity to high stress levels – and these often build on, reinforce, or influence each other. The diversity of drivers identified, the interplay among them, and research suggesting that families with a child in institutional care likely experience the negative condition for five to six of these drivers (Kraguli and Pop 2012) means that there are tens of thousands of possible combinations of drivers forming pathways to separation.

To understand where interventions might disrupt some of those pathways, we reviewed the available grey literature and limited number of relevant peer-reviewed studies that discuss factors that may help families stay together (Laumann 2015). These include a sense of

Figure 2 Concept map of drivers of family–child separation with illustrative family social strengthening (medium tint), economic strengthening (light tint), and combined (dark tint) intervention points indicated



Note BS – business skills coaching; CT – cash transfers; FL – financial literacy; HH – household; IGA – income-generating activities; LS – life skills; MSA – matched savings accounts; PS – parenting skills; SPM – enterprise selection, planning, and management; SW – social work; VSLA – village saving and lending associations.
Source Authors' own.

belonging or community acceptance and existence of social safety nets (4Children 2015), along with access to childcare (CCF Moldova and Hope and Homes for Children UK 2015), higher parental level of education (Akwara *et al.* 2010), parenting capacity (Annan *et al.* 2013), and positive emotional states (Roelen 2015). On the economic side, there are few studies that directly examine the link between economic activities and family preservation or retention in care, with those few focusing primarily on income support through government social protection programmes (Barrientos *et al.* 2013; Roelen and Kharki Chettri 2014; Roelen and Shelmerdine 2014). Common economic interventions in the field of child protection include limited consumption support (cash transfers), microloans or small grants, small business start-up kits, and vocational skills training and income-generating activities. There is also practitioner interest in using savings groups and sequenced 'graduation' approaches (Mattingly 2015; Chaffin and Kalyanpur 2014). However, a review by Ellis and Chaffin (2015) suggests that, as yet, there is no clear evidence indicating which economic strengthening interventions are effective in supporting family–child unity. We therefore mapped a number of economic and family

strengthening activities onto the concept map of drivers of family–child separation, to begin to build a theory of change indicating how and where we might expect interventions to affect outcomes (see Figure 2).

1.4 Main features of the projects

Against the backdrop of limited evidence and with a draft theory of change, we put out a call for proposals for projects that would integrate a core set of traditional social work and family strengthening activities (e.g. case management, counselling, parenting skills, community dialogues) with at least three commonly used or promising economic strengthening activities (e.g. cash transfers, financial literacy training, savings groups). We requested that proposed projects serve families at risk of separation (and devise a methodology to identify them), and support the reintegration of separated children into their families and communities, a process that continues past the time of a child's reunification with family.[7]

Following a competitive solicitation process, we selected two projects, both in Uganda.[8] Family Resilience (FaRe), led by the Association of Volunteers in International Service (AVSI) Foundation with partner Retrak, focused on (a) supporting the reintegration of street-connected children and children from a government remand home into family care, and (b) preventing family separation in at-risk households in urban/peri-urban slums thought to be hotspots for child separation. Economic Strengthening to Keep and Reintegrate Children in Families (ESFAM), implemented by ChildFund, focused on (a) supporting the reintegration of children who had been living in child care institutions and were returned to their families in response to a government deinstitutionalisation policy, and (b) preventing unnecessary family separation among at-risk households in districts with several orphanages. The identification and selection of households 'at risk' of separation differed slightly between FaRe and ESFAM, but each included a mix of quantitative and peer-informed (qualitative) vulnerability assessments that assessed both economic and child protection vulnerabilities. Households eligible for inclusion demonstrated need in both domains.

Both projects included case management processes focused on developing household plans, assessing child welfare, providing counselling, and referring to outside support and services as needed as part of family social strengthening activities. Some households also received parenting skills training or coaching, individually or in groups, and/or participated in community dialogues on issues related to family dynamics and relationships. These activities accompanied regular home visits that were intended to occur once per month to once per quarter, depending on the sub-population. The primary household economic strengthening activities across the projects included time-limited unconditional cash transfers, financial literacy training, matched savings accounts, and village saving and lending associations (VSLAs), with some apprenticeships, income-generation training, and child- and youth-savings groups.

2 Evaluation design considerations

The absence of a clear and simple $x \rightarrow y \rightarrow z$ causal pathway is perhaps not unusual for development programmes, but the added complexity of cross-sector factors affecting family–child separation generates a number of challenges for integrated development evaluation design. A full-factorial (2x2) randomised control trial (RCT) could help to isolate the effects of separate project elements and provide information on their interaction by allowing the comparison of 'family social strengthening-only' to 'family social strengthening activities plus (different) economic strengthening activities' to a control group that received only the standard of care.

We concluded, however, that important features supporting such a stringent study were not in place. First, we have limited evidence in support of any particular household economic strengthening interventions in this context to justify the selection of one economic intervention over another. We also did not know baseline rates of separation to inform estimations of statistical power. The numerous potential pathways to the primary outcome further complicated the comparison of the integrated intervention to any single intervention. And we also considered contextual factors, such as the geographic complexities of reintegrating children (research staff cannot control which children are ready for reunification, or when or where), and logistical factors, such as limited staff experience implementing a combined family and economic strengthening programme. Based on these considerations, along with time and resource constraints, we ruled out a factorial RCT design fairly quickly. Instead, our focus from the beginning was on building an evaluation framework that would generate robust data to address programmatic learning needs while filling in some of the evidence gaps.

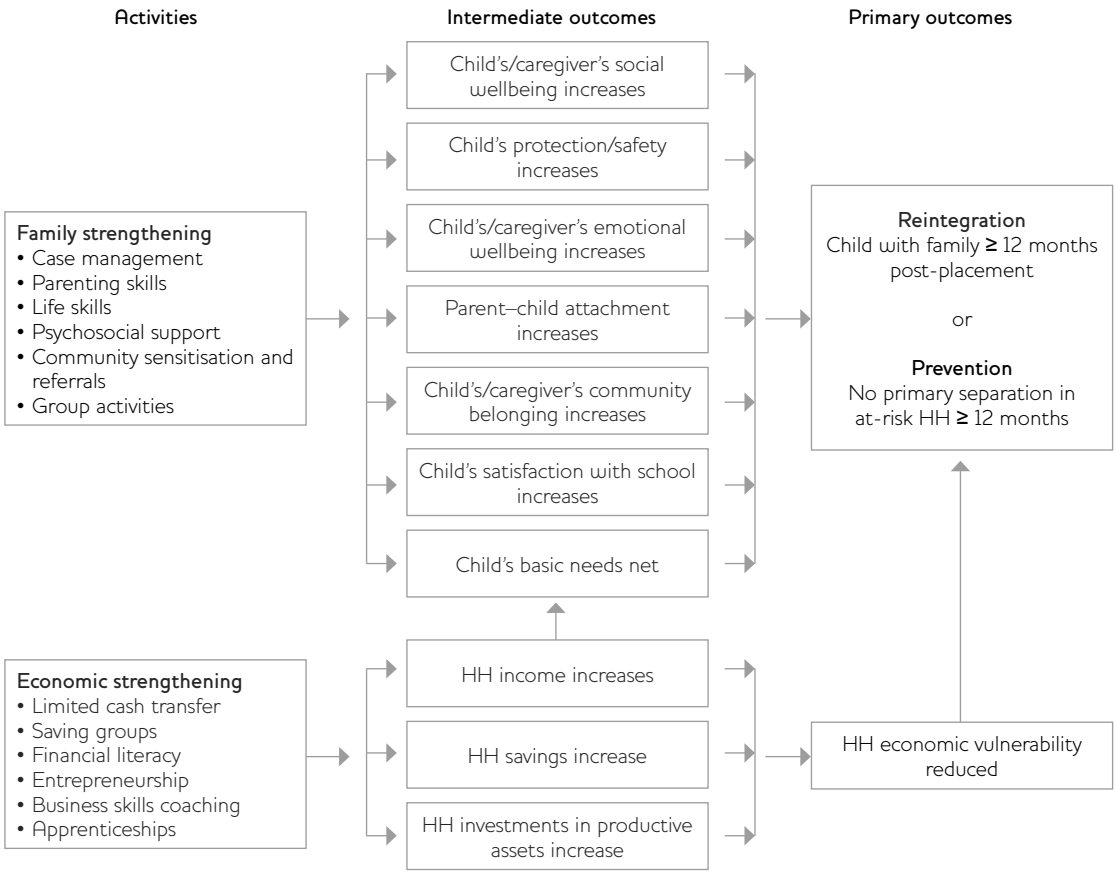
2.1 Evaluation scope and objectives

The scope of the Family Care project evaluations is largely defined by two related objectives.

> Objective 1: Demonstrate and assess the *effectiveness* of different household economic strengthening programmes in the context of family–child reintegration and prevention of family–child separation.

The focus of this objective is on evaluating how well different types of economic strengthening activities work when employed in support of family unity. We seek to understand the extent to which household economic strengthening activities provided in this context help to prevent family–child separation and produce positive changes in family social and economic indicators, and for which types of families. We recognise that our data, which only cover project participants, do not allow for an impact evaluation design to measure effectiveness as a net attributable impact. Instead, to assess effectiveness, we employ several methods to examine, rather than measure, the outcomes of the projects on their beneficiaries.

Figure 3 **Abbreviated theory of change illustrating activity inputs by domain and outcomes**



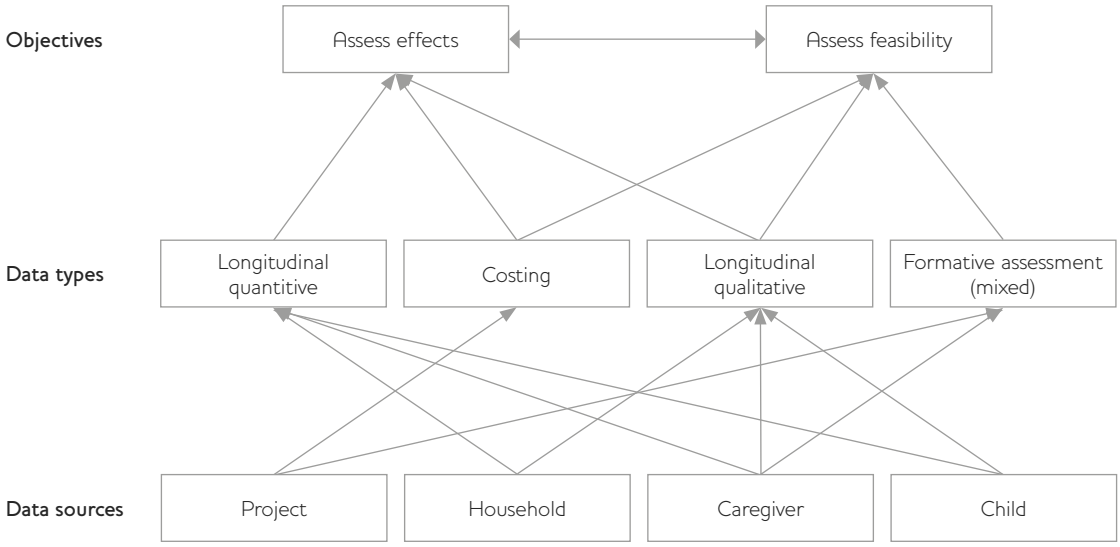
Note HH – household.
Source Authors' own.

Objective 2: Demonstrate and assess the *feasibility and appropriateness* of integrating targeted household economic strengthening interventions into programming in the context of family–child reintegration and prevention of family–child separation.

The intent of this objective is to generate learning from the implementation of the projects that will inform guidance on how reintegration and prevention-of-separation programmes can best incorporate economic strengthening interventions. We seek to understand the mechanics of *how* economic strengthening programmes work for families, including (a) the complementarities of the household economic strengthening activities and other family strengthening activities, (b) any areas where economic strengthening activities are disruptive to family strengthening goals, and (c) the comparative costs to provide economic strengthening programming within the reintegration and prevention contexts.

The scope of our programme evaluation therefore includes outcome evaluation questions and formative evaluation questions, with specific

Figure 4 **Diagram of evaluation design elements**



Source Authors' own.

intention to learn about activity integration and interaction effects from the answers to both sets of questions.

2.2 Outcomes

We used the concept map in Figure 2 along with the objectives above to help inform the selection of appropriate outcomes to reflect family social and economic strengthening. Selection of the outcomes – and development of the tools to measure them – was a multidisciplinary effort, reflecting input from experts in child protection and economic strengthening, local practitioners, local reintegrating families, and researchers. Given our objective to understand the effects of integrated family and economic strengthening on family–child separation, we designate permanency of placement for reintegrating children and continued family care for at-risk children as our primary outcome. We also consider this to be our 'synergistic' outcome as, according to the theory of change, successfully preventing separation or re-separation requires, in most cases, addressing drivers in both family and economic domains (see Figure 3). However, given the diversity of potential pathways to separation, and an unknown base rate of family–child separation, we also want to be sure to assess family social and economic outcomes separately. These intermediate outcomes also allow us to understand the relative importance of changes in different areas of family life.

3 Mixed methods evaluation design

The combination of outcome and formative evaluation questions in our study resulted in a multilevel mixed methods evaluation design. Our design comprises four elements based on four types of data – longitudinal quantitative data, costing data, qualitative longitudinal data, and activity-based process evaluations – each intended to contribute

specific types of information associated with our learning objectives (see Figure 4). We describe our evaluation framework as multilevel, in addition to mixed method, to describe the 'layering' of data sources and analysis methods within concurrent related-but-separate assessments. In this section, we present each element, the types of data collected, and how they address our integrated development objectives.

3.1 Evaluation population

We invited caregivers in all FaRe and ESFAM households (n=~1400) to participate in the evaluation research, along with one 'index' child from each household. For households reintegrating a child, the reunified child was the index child. For households at risk of separation, the index child was selected based on a quantitative assessment of relative vulnerability to separation; the child between the ages of seven and 17 with the highest vulnerability score was selected as the index child in these families.

3.2 Longitudinal quantitative data

We collected information from all participating households at three points over the 18-month implementation period, to gather longitudinal quantitative (closed-form) data about both economic and wellbeing outcomes. All quantitative data were collected in-person, using questionnaires administered by external research staff (ESFAM) or project staff (FaRe). The projects used a Family Care-developed instrument to collect information from caregivers on five domains of wellbeing (social, parent–child attachment, community belonging, emotional belonging, and child care and protection) and from children on six domains of wellbeing (enjoyment of education, social, parent–child attachment, community belonging, emotional belonging, and child protection). These domains track to family strengthening activities. The projects used a separate instrument to solicit household information from the caregiver on financial and economic indicators (e.g. income, savings, assets), food security, access to basic needs and education, and child protection issues (e.g. harsh discipline, child separation). The two projects used slightly different instruments to collect this economic strengthening information, but we harmonised the questions related to the main indicators.

We designed these baseline–midline–endline instruments to address Objective 1, the extent to which economic strengthening interventions provided in the context of family strengthening affect the outcomes of interest. With these observational data from a relatively large sample, we can examine many characteristics and outcomes related to our integrated development research questions, including: (a) rates of separation for both reintegrating and at-risk households, (b) changes in a family's wellbeing (from both caregiver and child perspective), (c) changes in a household's economic indicators and vulnerability, (d) characteristics of households that experience a separation compared to those that do not, (e) correlations between changes in family wellbeing or economic status and separation rates, and (f) how the types of changes observed in analyses a–c differ (or not) according to the type (and amount) of economic strengthening intervention received.

Analyses e and f address questions associated with the study's general objective to identify which economic strengthening activities might work for which types of families, necessary to inform the generation of guidelines for practitioners. The large sample size allows the use of difference-in-differences analysis and structural equation modelling to help us investigate sub-group variations and likely contributors to observed effects. For example, preliminary analyses suggest that at-risk households and reintegrating households that received a cash transfer were similarly vulnerable at baseline but experienced statistically significantly different changes between baseline and endline, with the at-risk households showing more improvement. While we do not have a comparison group to allow for measurement of an attributable net programme effect for either group of households, the cross-group statistical analysis using longitudinal quantitative data offers insights for programming consideration. The collection of midline data in addition to the before and after data allows us to better examine the trends within and across households over time, relative to both the family and economic strengthening outcomes. We can explore whether and when certain covariates impact the trajectory of household changes.

3.2.1 Set-theoretic analyses

Even with longitudinal data, what we cannot do in the absence of a comparison group to estimate a counterfactual is attribute observed changes to specific programme activities. However, there are other strategies we can employ to unpack the interactions and importance of project activities relative to the primary outcomes. In this evaluation, we use qualitative comparative analysis (QCA). QCA is a set-theoretic data analysis technique that relies on inferential logic or Boolean algebra, rather than statistical correlation, 'to specify the different combinations of conditions linked to the selected outcome, based on the features of the positive cases that consistently distinguish them from the negative cases' (Ragin n.d.: 11). QCA was developed in the field of political science within the discipline of case study research as a way to identify, among a small number of cases (usually countries), the policy conditions necessary and sufficient to generate a particular outcome.

Yet despite its name and its original purpose, QCA is neither limited to qualitative data nor small-n samples. Rather, it is a highly structured and quantitative – but not statistical – means of identifying patterns within a data set based on the creation of like sets. The UK Department for International Development (DFID) describes QCA as a 'newer method' of evaluation for which there is compelling rationale and scope (Stern *et al.* 2012: 49), and though QCA has not been used extensively in development evaluation, there are a handful of recent examples (Welle *et al.* 2015; Stedman-Bryce *et al*. 2015; Schatz and Welle 2016). Unlike linear regression-based approaches, set-theoretic techniques like QCA (and also coincidence analysis) do not try to organise all data points to one best-fit line, but rather they allow for multiple 'lines' of connection from inputs to outcomes, recognising that in complex systems, there is often more than one route from starting point to desired effect. This

perspective is exactly what integrated development is about – different activities/inputs and how they combine to generate effects – and consequently aligns with both our concept map and research objectives.

In our evaluation, we use QCA to examine how some project participants succeeded in keeping children in family care (positive cases) while others did not (negative cases), based on the combination of project activity inputs they received. The analysis of these positive and negative cases allows us to identify the necessary and sufficient conditions for pathways through which the ASPIRES Family Care project participants achieve prevention of family–child separation and successful reintegration of children into family care. And indeed, the QCA analyses help both to clarify the linkages in our concept map and to identify which combinations of project activities contribute to observed effects. The analyses can also inform aspects of Objective 2, namely how family strengthening activities might be bolstering (or detracting from) economic strengthening activities, or vice versa.

Though QCA does not require a large sample, we are using our endline quantitative data to populate the 'truth tables' that serve as input for the QCA analyses. A large sample provides greater variation between households within the sets of positive and negative cases, which is particularly helpful in the Family Care project evaluations, given the number of economic and family strengthening activity inputs. And, because the incidence of family–child separation is unknown, a larger sample provides a greater chance to observe the negative cases, i.e. where separation does occur. The larger sample also allows for disaggregation by type of household (at-risk/reintegrating) or geography (rural/urban) to address the question of which types of economic strengthening activities appear to help which types of households.

3.3 Costing data

In response to Objective 2, we also collected cost information on all project activities over the life of the project, to examine the cost of the integrated programme and the costs of the different interventions. While integrated programming could produce economies of scale, where an integrated programme costs less than two separate interventions, the complexity of an integrated programme could also cause costs to multiply. It is therefore as important to understand the effects of integrated programming on costs as on integration outcomes. We prepared a Microsoft Excel tracking workbook and written guidance to enable FaRe and ESFAM finance staff to capture costs in a way that allows the aggregation and disaggregation necessary to examine several formative evaluation questions (e.g. the average costs of family strengthening and economic strengthening activities independently and combined). Within our larger evaluation design, these data provide an opportunity to derive costs per beneficiary and, in conjunction with the quantitative outcome data, may provide suggestive information about cost-effectiveness.

In practice, the cost data proved the most challenging to collect. Despite their willingness and intent to provide quarterly costing updates, our partners found this activity to be very difficult as their financial systems did not capture information in ways that easily translated to the costing templates we provided. Rather, programme implementation, management, and finance staff often had to sit together and review physical documentation to produce the updates. We reduced our requirements for precision to make the exercise more manageable, acknowledging that less precision affects the quality of the data and the analyses we can perform. Nonetheless, the cost estimates for project elements are essential to the programmatic guidance we are generating for the integration of economic and family strengthening activities.

3.4 Longitudinal qualitative data

We also collected longitudinal qualitative data from a stratified purposive sample of 80 households, 20 each of reintegration households and at-risk households per project, stratified further by sub-county to ensure geographic variation. Households were randomly ordered within each sample stratum and then systematically selected in order according to the gender of the index child to maintain gender diversity. For each selected household, the study team conducted separate open-ended qualitative interviews with both a caregiver and the index child. We interviewed reunified families three times; families at risk of separation four times. All qualitative interviews were conducted by external research staff in pairs, who conducted the caregiver and child interviews separately and in private. All interviews began by building a guided 'life narrative' that provides a snapshot of life within the household on specific domains and captures changes in these domains over time.

In an integrated development context, qualitative data are extremely useful for identifying and analysing interactions and synergies. We asked direct questions about interactions, such as 'What connections do you see between the [family strengthening activity received] and the [economic strengthening intervention]?' And we see examples of complementary effects in both directions:

> *They support each other because when you learn how to save money and spend it well, definitely it will help you become a better parent because then you can provide for their needs better and even pay their school fees.* (Gulu at-risk household)[9]

> *The caregiver thinks that the two project activities are both important and if delivered well they can positively impact her household. For example, she thinks that if there's cooperation among family members they can achieve positive decision of labour where for instance as she attends to the garden, her daughter attends to the grocery stall. This would promote unity and improve on the household productivity*. (Luwero reintegration household)[10]

We also look for examples of interactions and synergies in participant narratives. The longitudinal nature of the open-ended interviews enabled us to develop household-level case histories that include

rich descriptions of changes in each family from caregiver and child perspectives, on both the economic and livelihoods situation and on the interpersonal dynamics of wellbeing. A caregiver may describe how her ability to generate income (an economic outcome) reduces her stress, which in turn allows her to parent better and improve the caregiver–child connection (a family wellbeing outcome), which she in turn credits for the lack of family–child separation in her household.

> *The children are happy because we are no longer bothered by the land lady. Even recently I told her that [we] had sent this month's instalment and she was happy. Our relationship is now better because we are no longer constrained by money problems. I am no longer worried as before, so I do not take out my stress on the children by shouting at them. I talk to them in case they have done something wrong*. (Kampala at-risk household)[11]

In addition to answering a range of programmatic questions on their own, the qualitative data provide crucial contextual information to facilitate interpretation of the quantitative findings. For instance, when we see in the preliminary quantitative data that cash transfer households with a reunified child increased their income but decreased their savings, on average, from baseline to endline, we can look to the qualitative data for explanations of why this may be the case.

3.5 Formative evaluations of specific interventions

The quantitative, costing, and qualitative data described above contribute information related to Objective 2 and the relationships between family social and economic strengthening project elements. However, questions around the mechanics of overall implementation and specific project components require dedicated inquiry to generate information on how the household economic interventions functioned under the specific contexts of ESFAM and FaRe, along with perspectives on what worked from an implementation and supply point of view and what could be improved in similar situations. To do this, we undertook two formative evaluations per project on common economic strengthening activities: one on cash transfer interventions and one on VSLA interventions. We treated these evaluations as an 'external consultancy', with the scope of work drafted by our research team, shared with FaRe and ESFAM for their client perspective, and then carried out by two ASPIRES team members with some knowledge of the projects but little prior contact with them. Our ASPIRES colleagues then conducted fieldwork for both evaluations concurrently, including interviews and focus group discussions with project staff, finance staff, local government, cash transfer recipients, and VSLA members.

These evaluations provide important data on the feasibility and acceptability of specific economic strengthening activities conducted with a family strengthening aim. For example, we learned about the appropriateness of the amount and duration of cash transfers (likely too little cash for too short a period for some families), the ability of very poor people to save (they can, and less-poor VSLA members are willing to accommodate them), and the workforce needs and challenges

in doing this type of integrated programming (there are many). In addition, the findings from these formative evaluations also help us to interpret the data on effects, in the same way the qualitative data do, but with more specificity on designated interventions. For example, a finding from the VSLA formative evaluation – that VSLA participants felt reduced social isolation and a sense of having a social safety net in the community – shows a complementary effect of the integrated programming, in that the participants' sense of community belonging and social wellbeing (two of the family strengthening objectives) are accomplished through an economic strengthening activity.

4 Discussion

In an evaluator's ideal world, development problems would come with a neat underlying cause–effect structure against which to design and test interventions. But the real world is messy. As Neil deGrasse Tyson said (2016), 'In science, when human behavior enters the equation, things go non-linear. That is why physics is easy and sociology is hard.' Integrated development is sociology – and anthropology, epidemiology, economics, social work, public health, diplomacy, and a dozen other disciplines working together. It is usually, as in the case of the Family Care project, closer to circular than linear, with multiple interacting feedback loop 'causes' for each effect. Response to this complexity – and evaluation of it – requires integrated and complementary solutions.

An obvious choice for evaluating a two-sector integrated development programme is a 2x2 factorial design RCT. The factorial design facilitates measurement of interaction and synergy effects and random assignment eliminates selection bias. A recent review found, however, that while 70 per cent of integrated development impact evaluations employed randomisation, only 7 per cent used a factorial design (Ahner-McHaffie *et al.* 2018). In our case, we did not have the resources to conduct an RCT, but as important, we concluded there was not enough evidence regarding the individual interventions or information on how they should be implemented and integrated to justify a factorial RCT. The 'gold standard' approach is less golden if the results are null because the interventions being tested are not well understood (in theory or practice) (Brown 2017).

In situations like this, where the focus is still on understanding how an integrated programme might work, rather than measuring net impacts, descriptive mixed method observational research plays a critical role in building the evidence base. Done well, it can identify trends, surface unintended or unexpected consequences of new integrated programming, provide explanations to help interpret those trends, and inform programme design and implementation decisions. Relatedly, in the absence of an impact evaluation (and even with it), qualitative data can provide critical explanatory power to a research design.

Ahner-McHaffie and colleagues concluded in their review that 'using a mixed method approach to include examination of cost efficiencies and qualitatively assessing synergies offer[s] value for determining how

integration factors in to the program findings' (2018: 10). Our multilevel mixed methods approach was perhaps unorthodox in its blend of formative and outcome evaluation objectives, yet because the elements were designed as complementary, they yielded a broad range of robust data to address those objectives. The longitudinal data – quantitative and qualitative – provide different perspectives from which to assess questions about both the process and effects of integration and synergy. The large and varied sample for the quantitative data also allows disaggregation for within-group comparisons, and flexibility to conduct QCA to help understand which of the many project elements are necessary for success. The formative evaluations, along with the costing data, help us to understand the feasibility of interventions being used in a new and different context and even reveal some complementary effects. In these ways, the Family Care evaluation data will substantially contribute to building the evidence base and providing programmatic guidance on the process of implementing economic strengthening activities in support of family unity, indicating directions for future integrated programming – and evaluation – in this area.

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.

† This project was carried out under United States Agency for International Development (USAID) Cooperative Agreement No. AID-OAA-LA-13-00001. The contents are the responsibility of FHI 360 and do not necessarily reflect the views of USAID or the United States government.

1 Corresponding author. FHI 360, Durham, North Carolina, USA. 359 Blackwell Street, Suite 200, Durham NC 27701, USA. enamey@fhi360.org.

2 FHI 360, Altadena, California, USA. 1825 Connecticut Avenue, NW, Suite 8000, Washington DC 20009, USA. llaumann@fhi360.org.

3 FHI 360, Washington, District of Columbia, USA. 1825 Connecticut Avenue, NW, Suite 8000, Washington DC 20009, USA. abrown@fhi360.org.

4 The Guidelines for the Alternative Care of Children by the United Nations General Assembly in 2009 affirm the 'family as the fundamental group of society and the natural environment for the growth, wellbeing and protection of children', children's right to 'live in a supportive, protective and nurturing environment', and the responsibility of the state to return children to family care, ensure that families have access to support as caregivers, and ensure appropriate alternative care if needed (United Nations General Assembly 2010: 2–3).

5 www.fhi360.org.

6 The Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) project, supported by the President's Emergency Plan for AIDS Relief (PEPFAR) and USAID and managed by FHI 360, supports gender-sensitive programming, research, and learning to improve the economic security of highly vulnerable individuals, families, and children. The Family Care project under ASPIRES, funded by the Displaced Children and Orphans Fund of USAID, developed the work described here.
7 The *Guidelines on Children's Reintegration* define reintegration as the '… process of a separated child making what is anticipated to be a permanent transition back to his or her family and community (usually of origin), in order to receive protection and care and to find a sense of belonging and purpose in all spheres of life' (Interagency Group on Children's Reintegration 2016: 1).
8 The initial intention had been to select projects in two different regions (eligible countries included Rwanda, Uganda, Colombia, Guatemala, and Haiti), but the projects that scored highest in the competitive process were both proposed for implementation in Uganda.
9 Interview conducted July 2017.
10 Interview conducted July 2017.
11 Interview conducted July 2017.

References

4Children (2015) *Expanding the Evidence Base to Inform Next Generation Programming: Keeping Children in Healthy and Protective Families – Research Inception Meeting Synthesis Report*, Baltimore MD: Catholic Relief Services

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 1 October 2018)

Akwara, P.A. *et al.* (2010) 'Who is the Vulnerable Child? Using Survey Data to Identify Children at Risk in the Era of HIV and AIDS', *AIDS Care* 22.9: 1066–85

Annan, J.; Bundervoet, T.; Seban, J. and Costigan, J. (2013) *A Randomized Impact Evaluation of Village Savings and Loans Associations and Family-Based Interventions in Burundi*, New York NY: International Rescue Committee

Bakker, C.; Elings-Pels, M. and Reis, M. (2009) *The Impact of Migration on Children in the Caribbean*, New York NY: UNICEF

Barrientos, A.; Byrne, J.; Villa, J.M. and Peña, P. (2013) *Social Transfers and Child Protection*, Innocenti Working Papers 2013-05, Florence: UNICEF Office of Research

Berens, A.E. and Nelson, C.A. (2015) 'The Science of Early Adversity: Is There a Role for Large Institutions in the Care of Vulnerable Children?', *The Lancet* 386.9991: 388–98, https://doi.org/10.1016/S0140-6736(14)61131-4 (accessed 14 September 2018)

Boothby, N. *et al*. (2012) 'What are the Most Effective Early Response Strategies and Interventions to Assess and Address the Immediate Needs of Children Outside of Family Care?', *Child Abuse and Neglect* 36.10: 711–21

Brown, A.N. (2017) *Null Results Should Produce Answers, Not Excuses*, R&E Search for Evidence, Washington DC: FHI 360, http://researchforevidence.fhi360.org/null-results-should-produce-answers-not-excuses (accessed 1 October 2018)

Browne, K. (2009) *The Risk of Harm to Young Children in Institutional Care*, London: Save the Children

CCF Moldova and Hope and Homes for Children UK (2015) *Preventing Family Separation/Supporting the Tracing, Reunification and Reintegration of Separated Children into their Families and Communities*, Chișinău and Salisbury: CCF Moldova and HHC UK

Chaffin, J. and Kalyanpur, A. (2014) *What Do We Know about Economic Strengthening for Family Reintegration of Separated Children?* New York NY: CPC Network

Csáky, C. (2009) *Keeping Children Out of Harmful Institutions: Why We Should be Investing in Family-Based Care*, London: Save the Children

deGrasse Tyson, N. [@neiltyson] (2016) 'In Science, When Human Behavior Enters the Equation, Things Go Nonlinear. That's Why Physics is Easy and Sociology is Hard', tweet, https://twitter.com/neiltyson/status/695759776752496640?lang=en (accessed 26 September 2018)

Delap, E. (2013) *Towards a Family for Every Child: A Conceptual Framework*, London: Family for Every Child

Dozier, M.; Zeanah, C.H.; Wallin, A.R. and Shauffer, C. (2012) 'Institutional Care for Young Children: Review of Literature and Policy Implications', *Social Issues and Policy Review* 6.1: 1–25

Ellis, C.M. and Chaffin, J. (2015) 'Evaluations of Outcomes for Children and Youth from NGO-Supported Microeconomic Interventions: A Research Synthesis', *Enterprise Development and Microfinance* 26.2: 104–21, https://doi.org/10.3362/1755-1986.2015.012 (accessed 14 September 2018)

EveryChild (2009) *Every Child Deserves a Family: EveryChild's Approach to Children without Parental Care*, London: EveryChild, https://bettercarenetwork.org/library/social-welfare-systems/standards-of-care/every-child-deserves-a-family-everychilds-approach-to-children-without-parental-care (accessed 1 October 2018)

Interagency Group on Children's Reintegration (2016) *Guidelines on Children's Reintegration*, www.familyforeverychild.org/wp-content/uploads/2016/08/RG_Digital_DC.pdf (accessed 25 July 2018)

Kraguli, J. and Pop, D. (2012) *Preventing the Separation of Children from their Families in Bosnia and Herzegovina: Review of Hope and Homes for Children ACTIVE Family Support Programme in Bosnia and Herzegovina 2003–2010*, Salisbury: Hope and Homes for Children

Laumann, L. (2015) *Household Economic Strengthening in Support of Prevention of Family–Child Separation and Children's Reintegration in Family Care*, Washington DC: FHI 360

Mattingly, S. (2015) *Symposium Report: Keeping Children and Families Together with Economic Strengthening*, www.marketlinks.org/sites/marketlinks.org/files/resource/files/STRIVE_Symposium_report-final_0.pdf (accessed 1 October 2018)

Petrowski, N.; Cappa, C. and Gross, P. (2017) 'Estimating the Number of Children in Formal Alternative Care: Challenges and Results', *Child Abuse and Neglect* 70: 388–98, https://doi.org/10.1016/j.chiabu.2016.11.026 (accessed 14 September 2018)

Ragin, C.C. (n.d.) *What is Qualitative Comparative Analysis (QCA)?*, http://eprints.ncrm.ac.uk/250/1/What_is_QCA.pdf (accessed 26 September 2018)

Roelen, K. (2015) 'Economic Strengthening and Child Wellbeing and Care: Positive Effects and Adverse Consequences', paper presented at the STRIVE Symposium, Washington DC, 6 March, www.cpcnetwork.org/wp-content/uploads/2015/04/March_6_2015_ES_Roelen-K-STRIVE.pdf (accessed 1 October 2018)

Roelen, K. and Kharki Chettri, H. (2014) *Researching the Linkages between Social Protection and Children's Care in Ghana: LEAP and its Effects on Child Well-Being, Care, and Family Cohesion*, London, Brighton and Accra: Family for Every Child, IDS and Challenging Heights

Roelen, K. and Shelmerdine, H. (2014) *Researching the Linkages between Social Protection and Children's Care in Rwanda: The VUP and its Effects on Child Well-Being, Care and Family Reunification*, London: Family for Every Child

Schatz, F. and Welle, K. (2016) *Qualitative Comparative Analysis: A Valuable Approach to Add to the Evaluator's 'Toolbox'? Lessons from Recent Applications*, Centre for Development Impact (CDI) Practice Paper 13, Brighton: IDS

Stedman-Bryce, G.; Schatz, F.; Hodgkin, C. and Balogun, P. (2015) *Medicines Transparency Alliance (MeTA) Evaluation: Testing MeTA's Underlying Intervention Logic*, London: e-Pact

Stern, E. *et al*. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, London: Department for International Development

Tobis, D. (2000) *Moving from Residential Institutions to Community-Based Social Services in Central and Eastern Europe and the Former Soviet Union*, A World Free of Poverty Series, Washington DC: World Bank

United Nations General Assembly (2010) *Guidelines for the Alternative Care of Children*, Resolution Adopted by the General Assembly, A/RES/64/142, New York NY: United Nations, https://resourcecentre.savethechildren.net/node/5416/pdf/5416.pdf (accessed 26 September 2018)

Wedge, J.; Krumholz, A. and Jones, L. (2013) *Reaching for Home: Global Learning on Family Reintegration in Low and Lower-Middle Income Countries*, New York NY: Better Care Network

Welle, K.; Williams, J.; Pearce, J. and Befani, B. (2015) *Testing the Waters: A Qualitative Comparative Analysis of the Factors Affecting Success in Rendering Water Services Sustainable Based on ICT Reporting*, Brighton: IDS and WaterAid

Applying Factorial Designs to Disentangle the Effects of Integrated Development*†

Holly M. Burke,[1] Mario Chen[2] and Annette N. Brown[3]

Abstract In this article, we discuss the study design and lessons learned from a full-factorial randomised controlled study conducted with beneficiaries of a youth programme in Pretoria, South Africa. The study assesses whether the integration of an economic strengthening intervention with an HIV-prevention education intervention improves economic and health outcomes beyond singular interventions. The selected youth were randomised into four groups: combined economic strengthening and HIV-prevention interventions; economic strengthening intervention only; HIV-prevention education intervention only; or no interventions. We conducted a pre-intervention and two post-intervention assessments with the participants to measure outcomes, including the primary outcome – prevalence of sexually transmitted infections. We discuss our rationale for the study design and the challenges faced when implementing it. We consider how features of the integrated programme, such as how synergy is assessed, and features of context, for example available sample size, determine which methods can be used to test the effectiveness of integrated programming.

Keywords: integrated; development; multidisciplinary; multisector; evaluation; synergy; interaction effects; HIV prevention; economic strengthening.

1 Background

Globally, an estimated one third of all new HIV infections occurs among youth aged 15–24, highlighting the importance of an HIV response targeting youth (UNICEF 2013). Evidence shows that girls who engage in intergenerational and transactional sex are especially vulnerable to HIV (Luke 2005; Leclerc-Madlala 2008). Several studies show that HIV prevention education can educate and build skills, which leads to safer sex practices and lower rates of HIV and other sexually transmitted infections (STIs) (Jewkes *et al*. 2006; Kirby, Laris and Rolleri 2007; Wingood *et al.* 2007). There is also evidence that

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.165
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'**; the Introduction is also recommended reading.

economic strengthening (ES) interventions can foster greater financial independence, which may reduce the incentive to participate in transactional and intergenerational sex, and increase bargaining power in sexual relationships, for example by insisting on condoms (Swann 2018). Swann (2018) reviews a large body of evidence and concludes that cash transfers and educational support are effective in reducing self-reported HIV risk behaviours, especially among adolescents.

However, clinical evidence supporting these behavioural findings is currently lacking, and evidence for other types of economic strengthening interventions that may be more sustainable, such as savings and financial education, is even less conclusive (*ibid*.). Furthermore, HIV prevention and economic strengthening interventions are often conducted in isolation, despite HIV having both economic and social drivers. Some research has yielded promising results suggesting that interventions with economic and social components build skills to improve financial wellbeing, women's empowerment, and reduce intimate partner violence, thus reducing the vulnerabilities of HIV (Pronyk *et al*. 2006; Kim *et al*. 2007; Gupta *et al*. 2013). Kim *et al.* (2009) conducted a cluster randomised trial of a gender and HIV training programme combined with a microfinance intervention implemented in rural South African villages, and found greater reductions in women's HIV risk behaviours compared to groups receiving only the microfinance intervention and no intervention. However, few studies have investigated whether there is synergy in multisectoral integration; that is, whether the integration of two interventions produces a combined effect greater than the sum of the interventions' separate effects (Ahner-McHaffie *et al*. 2018).

To build the evidence base around the efficacy of integrated interventions for HIV prevention, we conducted a full factorial randomised controlled study to assess whether the integration of an economic strengthening intervention with an HIV prevention education intervention improves health and economic outcomes for adolescents beyond singular interventions. In this article, we discuss our rationale for the study design and the challenges we faced when implementing it. We describe our analysis plan as designed and how it can and cannot be adjusted to account for the implementation challenges. We discuss how features of the integrated programme, such as how synergy is assessed, and how features of the context, such as the available sample size, determine which methods can be used to test the effectiveness of an integrated model within a full factorial design.

2 Study overview

2.1 Study participants

The study was conducted with the adolescent beneficiaries of a local, not-for-profit organisation implementing a programme providing support to poor families affected by HIV in four communities in Gauteng Province, South Africa.

2.2 Intervention description

FHI 360's Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES)[4] project provided technical assistance to the programme implementer to implement the economic strengthening and HIV prevention interventions, which were both educational interventions. The economic strengthening intervention – Impumelelo – builds on the Life Poa curriculum developed by YouthSave Kenya, and covered the topics of budgeting and saving, education on different savings options, and introduced the topic of earning money. The HIV prevention intervention, an updated version of the existing Vhutshilo curriculum that has been used with vulnerable youth throughout South Africa, covered expressing feelings, dealing with loss and grief, decision-making, drugs and alcohol, HIV and other STIs, healthy relationships, safer sex, and contraception. Each intervention consisted of 16 in-person sessions with a group of approximately 15 youth that lasted approximately 90 minutes. Participants in the combined intervention group received the two interventions sequentially (32 sessions total), though the order of the interventions varied due to programme staffing constraints.

2.3 Study design

From January to July 2016, 1,773 females and males aged 14–17 years were randomised in a 1:1:1:1 ratio to the four study groups: ES and HIV (combined economic strengthening and HIV prevention education interventions), ES (economic strengthening intervention only), HIV (HIV prevention education intervention only), or control (no additional education interventions). All participants received the programme's standard package of services which includes food and education support and linking families with community services, such as access to government grants. This trial is registered with ClinicalTrials.gov, number NCT 02888678.

2.4 Data collection

Employing a panel design, we collected data from the same participants three times during the study: at baseline (before the interventions began) and at two endlines. Endline 1 occurred shortly after the interventions finished and endline 2 occurred approximately eight months later to investigate the sustainability of the treatment effect. During each round of data collection, participants completed an in-person behavioural survey using audio computer-assisted self-interview (ACASI) technology where participants can read questions on a computer screen, hear the questions read to them privately through headphones, and indicate their responses without anyone viewing their selections during the interview. Participants also provided a urine sample for STI and pregnancy testing. Participants with positive STI or pregnancy tests were notified and referred for treatment or services after each round of data collection. Data collection ended May 2018 and analysis is ongoing.

We documented the implementation processes and costs associated with implementing the interventions. Programme staff, with support

from research staff, documented the resources used to implement the interventions at the programme level using electronic spreadsheets specifically tailored for the study. We also interviewed 22 purposively selected programme staff to gather information about the inputs needed to implement the interventions and their perspectives on the implementation of the interventions, including challenges experienced and resolutions to those challenges. Interviewees included the project coordinator, master trainer, programme managers, and facilitators of the economic strengthening intervention and the HIV prevention intervention.

2.5 Outcomes

The primary outcome for testing the effectiveness of the integration model is STI prevalence, defined as a positive test result for gonorrhoea, trichomoniasis, or chlamydia infection. We selected these non-viral STIs because they are common and treatable, and can be tested in urine samples obtained from both males and females. The secondary outcomes (all self-reported, except pregnancy) are:

1 Pregnancy (performed on the urine specimens of female participants);

2 Engaging in protective sexual behaviour, defined as self-reported abstinence or consistent condom use over the past six months;

3 Engaging in transactional sex in the past six months;

4 Having two or more sexual partners in the past six months;

5 HIV knowledge;

6 Financial literacy;

7 Participation in a savings group;

8 Opening a savings account;

9 Net change in savings in past year;

10 Saving for education;

11 Caregiver being primary provider of money to youth for savings;

12 Participation in household budgeting.

3 Challenges and approaches to evaluating an integrated programme

Evaluations of complex programmes, such as those with integrated multisectoral interventions, face important challenges. In this section, we discuss the challenges we faced and the decisions we made when designing and implementing the evaluation of a programme integrating interventions from the health and economic sectors. We specifically discuss how synergy is assessed, sample size considerations, and our analysis strategy and outcomes.

3.1 Challenge 1: how to define and measure integration and synergy effects

To assess the effect of an integrated programme against single interventions, we need to compare the integrated programme with each intervention implemented separately. A control group with no intervention or standard of care is also needed to determine the net effect of each single intervention and to determine how much better (or worse) the integrated programme is in affecting the outcomes. A systematic review found, however, that most experimental evaluations of integrated development programmes are 'two-arm' studies, comparing a group participating in the integrated programme to a control group not participating in the programme (FHI 360 2014). The review concluded that these comparisons preclude any assessment of whether single interventions achieve similar results as the integrated programme or what effects are attributable specifically to the integration.

A second systematic review, focused more specifically on this concern, assessed whether studies evaluating integrated development programmes measured synergistic effects (Ahner-McHaffie *et al.* 2018). Two programmes are said to work synergistically if the effects of the integrated intervention are amplified beyond the sum of the effects of each single sector intervention. Among the 601 impact evaluations included in this second review, 12 used partial factorial designs, and 26 used full factorial designs. In a full factorial design (a '2x2 design', assuming the integrated programme combines two interventions), the evaluation analyses data across four arms (or participants' groups), including separate arms for each of the interventions alone, for the programme that integrates those interventions, as well as for a control group. Only those impact evaluations with full factorial designs allow the measurement of the impact from integration and from synergy. The review finds, however, that most of the full factorial studies do not clearly discuss the distinct effects of synergy.

In our study, we used a full 2x2 factorial design and randomly assigned participants to ES, HIV, ES + HIV, or control. The synergy question is whether the whole is greater than the sum of the parts and can be stated as: is $1 + 1 > 2$? In a straightforward linear model, we can estimate the effect of each intervention, as well as the effect of the integrated approach using interaction terms to assess the effects of different interventions, whether implemented singly or in combination. We can consider the different effect scenarios as:

$1 + 1 = 2$, there is no synergy effect;

$1 + 1 > 2$, there is synergy (amplifying effects);

$1 + 1 < 2$, there is a detrimental effect.

However, even in the presence of detrimental effects (i.e. integrated programme not achieving the full sum of the single intervention effects), the integrated programme may still be considered beneficial if it improves outcomes more than each of the single interventions.

Put simply, it may still be the case that the whole is greater than either of the parts separately. Therefore, we posit another effect of interest that is less stringent for determining the value of the integrated programme: $1 + 1 > 1$. That is, we posit that integration produces a positive effect on top of the single intervention effect, even if the integration does not produce synergistic effects. This equation answers the question of whether the integrated ES and HIV intervention improves outcomes beyond what could be achieved by implementing either of the single interventions alone. It tells us if the integrated programme is the most effective of the three possibilities.

For our study, we based the sample size calculations on detecting $1 + 1 > 1$ to focus on the effects of the integrated programme and to mitigate the demand on sample size for adequately assessing the synergy hypothesis (see Section 3.2). To be exact, we focused on testing the two-sided version $1 + 1 \neq 1$ to allow for the possibility that integrating the programmes undermines the effect that could be achieved if we keep the interventions separated. This undermining effect of the integration could happen, for example, if programme staff or youth become overwhelmed by having too much to do in the integrated programme and therefore underperform in both components.

3.2 Challenge 2: the need for a large sample

Integrated development evaluations using a factorial design create several challenges for sample size. The first is that multi-arm studies divide the total sample into more groups than a two-arm (programme and control) study. As noted in Section 2.3, we divided our sample of 1,773 participants into four groups. If we think of simply testing each treatment arm against the control, we are only using half of the total sample for each test. We would need a factorial design sample size twice as large as the two-arm study to get the same power to measure the effect of the integrated programme against the control.

Another challenge comes from the potential of an interaction effect between the two individual interventions, where a positive interaction effect indicates synergy, one of the hypotheses we would like to test whenever possible. The challenge comes from the need for a larger sample size to detect interaction effects. Wolbers *et al.* (2011) provide sample size requirements for different levels of interaction effects. For example, they found that even under large interaction effects, doubling (strong synergy) or nullifying (zero effect of the integration) the effects of the single interventions requires fourfold the sample size of a two-group study (*ibid.*).

Unfortunately, we were not able to draw a sample large enough to test – with sufficient power – for interaction effects, at least not based on the assumptions in our power calculations. To address this, our primary hypothesis testing strategy will simply compare the integrated programme to each of the interventions separately. Thus, to conclude that the ES + HIV programme is effective, we will test

whether the integrated programme is statistically significantly more effective than each one of the interventions implemented separately. That is, ES + HIV compared to ES and ES + HIV compared to HIV. A statistically significant positive result in both comparisons will indicate a benefit associated with the integration model over what could be achieved with either of the interventions implemented separately. A statistically significant negative result in both tests will indicate a harmful effect of the integration. Effects in different directions may also indicate integration failure. We will also use the full data set to test whether there is a synergistic effect of the integrated programme (i.e. positive interaction effect), but understanding that we are likely to be underpowered to detect this effect.

Another challenge to sample size for evaluating integrated development programmes is multiple outcome testing, also called multiple comparisons or multiple inferences. The factorial design introduces multiple comparisons just based on the design alone. But even without a factorial design, evaluations of integrated programmes are likely to include measurements of many outcomes. Multiple outcome measurements arise from the desire to assess outcomes directly related to individual interventions and possible additional outcomes from the integrated programme. In Section 2.5, we present one primary and 12 secondary outcomes for our study.

The challenge arises because the more outcomes you test with the same data, the more likely you are to find statistically significant results for one or a few outcomes by chance alone. In statistical terms, the multiple comparisons problem leads to type I error inflation; you are more likely to reject a null hypothesis that is true (find an effect that is not there). The solutions to this problem require a more complex analytical model. For the more complex model to have the same power as a single comparison study, you need a larger sample size.

Based on the sample size we have, our strategy for addressing the multiple comparisons challenge is to pre-specify a primary outcome that we will use for our causal inferences. As noted above, this outcome is STI prevalence at each endline (with all positives at baseline and endline 1 receiving treatment so that prevalence starts at zero). We will present the analysis of the other 12 outcome variables as exploratory.

One more design consideration affecting sample size and statistical power for integrated development evaluations is that complex programmes are often implemented in groups, such as schools or clinics. If the randomisation is at the group or cluster level, then a larger sample is needed. Clustered designs are statistically less efficient because units within clusters are expected to be more homogeneous than they would be across clusters.

We were fortunate in our study to avoid a cluster randomised design. For our interventions, individual randomisation makes sense. The

interventions are designed to change individuals' behaviours, as opposed to programmes designed to change outcomes at a group level (for example, improving school quality or agricultural markets). The intervention activities (sessions), however, were designed to be carried out in groups. The programme implementer had no pre-existing groups that could be used for delivering the study interventions, so these needed to be formed for our study. Working closely with the programme implementer, we were able to recruit participants into our study, and randomise them into study arms, before they were then put into groups formed within each arm to receive the interventions (or not, in the case of the control arm). However, we recognise that because the intervention was delivered at the group level, group level differences may arise (for example, if some programme staff are better at delivering the interventions than others). If these group effects exist, we will control for them in our model and thus may lose some power for detecting the effects of interest. It is important to note that if we do find that the impact of the integrated programme is highly dependent on the facilitators' performance, it will imply that the programme is less scalable and potentially less useful for preventing HIV.

3.3 Challenge 3: implementation

A 2x2 factorial design requires three different implementations plus additional recruitment and data collection for the control. That means that the typical challenges studies face due to implementation are multiplied. Our study provides several examples. First, the need for the largest possible sample caused our timeline to be delayed because of the time it took to enrol nearly 1,800 eligible youth. We also faced delays from the need to hire the requisite staff to implement the two completely different and time-intensive education interventions. The implementer had to run three programme cycles to serve all the youth assigned to the intervention arms. The need for this repetition was driven in large part by the ES and HIV group, which required 32 separate training sessions.

Second, instead of the typical research format where the baseline and endline assessments are conducted for all participants at the same time and the interventions are implemented in between, we needed to enrol participants and collect the data on a continuous basis to complete the study on time. In addition to increasing the research costs, this meant that the first half of participants started interventions while we were still enrolling the second half of participants into the study. Moreover, at the end of the second intervention cycle, we began data collection for the first endline with the first group of enrolled participants (since they finished their interventions). This meant that some of the participants waited months between enrolment and their intervention to begin, while others started their intervention right after enrolment. These delays in a situation of multiple education interventions could result in contamination across intervention arms, as participants who know they will be taking a course but have to wait may seek discussions with others who are already in courses, regardless of whether it is the same course (i.e. study arm). Multiple intervention cycles also meant that

implementation could change over time, for example, as programme staff become more familiar with the curricula and become more experienced facilitators.

The third challenge to our study design from implementation constraints is that the implementer used a separate set of staff members to deliver the economic strengthening intervention from those who delivered the HIV prevention intervention. This specialisation facilitated the management of the work load and increased the quality and uniformity of delivery of the interventions. The HIV prevention intervention requires facilitators with higher skill levels, however, because they need to facilitate group sessions on the topics of HIV and sexual behaviour, which are more sensitive and stigmatised than topics like financial literacy and savings. In our 2x2 design then, the comparison of the two standalone intervention arms could be confounded or moderated by the quality of the facilitators.

3.4 Challenge 4: measuring outcomes

To fully evaluate this integrated programme, we needed to measure both economic and health outcomes. This meant we needed to collect a lot of data because these two sectors use very different indicators, methods, and timelines to measure outcomes. We settled on 13 outcome indicators, six health indicators, and seven economic indicators. To address multiple comparison concerns as mentioned above, we chose a primary outcome, STI prevalence, because it is a marker of unprotected sex, which is also the main risk factor for HIV transmission in the study setting. This clinical outcome is also considered less biased than self-reported measures and can be reliably measured in both boys and girls.

ES and HIV risk behaviours require different measurement techniques. Sexual behaviour that puts people at risk of acquiring HIV, such as engagement in transactional sex, is challenging to accurately measure through self-report because it is stigmatising, and sex work is illegal in many contexts. This required us to utilise additional (and often costly) technology to reduce reporting bias. In our study, we used ACASI techniques coupled with testing biological specimens for STIs and pregnancy. Economic strengthening outcomes, on the other hand, are less stigmatised and therefore may be more readily obtained through self-report.

ES and HIV risk behaviours may also develop differently over time, and this required us to take more than one endline measurement. For example, after participating in the ES intervention, youth may start to save money; however, it will take most youth a long time to save enough money to obtain higher education or skills training, start a business, or acquire enough productive assets to become financially independent to the point that they no longer need to engage in transactional or intergenerational sex to meet their needs. Whereas after participating in the HIV prevention intervention, youth may be more likely to engage in protective behaviours such as using condoms, with prevention messages

fresh in their minds. As time passes from the last lesson, we expect that sexual risk-taking will increase as the messages are forgotten or as other needs and priorities become more immediate, and as the youth mature into young adults. To investigate these varying timelines, we collected endline data twice: once right after the intervention ended and again as far out as our grant allowed. Three data points will also allow us to explore trends and the sustainability of effects over time.

3.5 Challenge 5: adherence and loss to follow-up

Adherence is a challenge for most interventions, not least of which are interventions that involve youth attending multiple group sessions after school. The challenge is greater for integrated programmes like ours where the integration is additive, because the integrated group has more to adhere to compared to the single intervention groups. It is also possible that one intervention type may have higher adherence than another because it has fewer requirements or is more desirable to the target population.

While participant retention is critical in the evaluation of any intervention, a time-intensive integrated intervention, like the one we evaluated, has the potential to result in differential loss to follow-up if participants in the integrated group are more likely to drop out of the study compared to other groups. Fortunately, through the diligent work of our research staff and support from the programme implementer, we had high overall participant retention throughout our study: 88 per cent at the first endline and 86 per cent at the second endline. We have not examined retention for each of our groups as of writing this article, but we are not expecting differential loss to follow-up, given our high overall retention. Low, non-differential loss to follow-up will reduce bias in our findings and gives us the best chance of reliably testing our research hypotheses.

4 Discussion

Integrated programmes that include economic strengthening components are increasingly being implemented to prevent HIV in resource-limited settings, but without rigorous evidence supporting this approach. We implemented a full factorial randomised controlled study to build the evidence base around the efficacy of integrated programmes for HIV prevention. While our study, like most evaluations conducted in development settings, faced financial and logistical constraints that prevented us from gathering a larger sample, we still believe the factorial design was the right decision. Unfortunately, we have not found any factorial design studies of integrated programmes that were able to draw sufficient samples to fully address all the challenges, but we hope that more integrated development programmes will be evaluated using factorial designs in the future. Given the statistical challenges, however, the results from these studies should be carefully interpreted. P-values can be easily misinterpreted if they are not clearly linked to the specific effects that are associated with hypotheses of interest in the evaluation of integrated programmes.

In designing this 2x2 factorial study of an integrated development programme, we felt that assessing synergy, sample size, outcome measurement, adherence, and loss to follow-up would be our main challenges. As discussed, we were not able to fully resolve the sample size challenges, which arise from multiple arms and multiple outcomes. Also, testing for synergy requires larger sample sizes. Our study protocol describes our analytical approaches for addressing these challenges in more detail. One approach we recommend is to focus the evaluation of integrated programmes on determining whether integration provides benefits beyond single interventions even if not amplifying the effects as expected under synergistic effects. Although ideal, synergy (i.e. the whole being greater than the sum of its parts) should not necessarily be the goal of every integrated programme. In terms of outcome measurement, we prioritised, at great expense, extensive data collection, including clinical indicators for STIs. Through careful implementation, we avoided major problems from loss to follow-up. What we did not fully anticipate were challenges to the study from the implementation constraints, such as differences among facilitators, for a multi-intervention, complex programme. In our analysis, we will pay particular attention to whether any of those challenges ultimately bias or contaminate our results.

Our main recommendation for future evaluations employing the full factorial design is to carefully consider all possible options to maximise the sample size. One approach to free up resources for a larger sample size is to not collect baseline data, as randomised controlled trials do not require baseline measures to help control for bias. That would not have worked in our situation, as we needed to treat all the baseline STIs to be able to measure STI prevalence as our outcome for endlines 1 and 2. It could work in other situations, though. Also, depending on the outcome targeted, researchers might consider collecting data for only one endline. In our case, we had two very different interventions that are on different timelines, so we saw value in collecting data for more than one endline, but programmes that integrate interventions with similar theories of change or similar timelines, especially interventions from the same sector, may not have this need.

While we were able to collect extensive data on a variety of outcomes for our full sample to conduct quantitative analysis, we recognise that mixed methods designs allow better analysis of the hows and whys, particularly for understanding synergistic effects. Unfortunately, funding cuts and unanticipated, increased expenses due to timeline delays forced us to eliminate our qualitative research component from the evaluation. Our plan was to conduct in-depth interviews with programme staff, youth participants, and their caregivers to describe whether, how, and why the interventions were perceived as effective in addressing economic and health outcomes among youth. Despite losing this important research component, the structured interviews we conducted with staff about their experience implementing the interventions gave us insight into how the interventions worked from the perspective of the implementers, and will contribute to the interpretation of the final

results. However, our recommendation is to include robust qualitative components in future integrated development evaluations.

We hope that the discussion of the challenges we encountered and our approaches for mitigating them contributes to the ongoing discussion on how integrated programmes should be evaluated and findings interpreted. Framing the interpretation of the results on the estimable effects in the context of study limitations will be important for the appropriate use of the findings.

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.

† The authors wish to acknowledge the following FHI 360 colleagues for their helpful review of this commentary: Mandy Swann, Emily Namey, and Michael Ferguson.

1 Corresponding author. FHI 360, Durham, North Carolina, USA. 359 Blackwell Street, Suite 200, Durham NC 27701, USA. hburke@fhi360.org.

2 FHI 360, Durham, North Carolina, USA. 359 Blackwell Street, Suite 200, Durham NC 27701, USA. mchen@fhi360.org.

3 FHI 360, Washington, District of Columbia, USA. 1825 Connecticut Avenue, NW, Suite 8000, Washington DC 20009, USA. abrown@fhi360.org.

4 The Accelerating Strategies for Practical Innovation and Research in Economic Strengthening (ASPIRES) project, supported by the President's Emergency Plan for AIDS Relief (PEPFAR) and the United States Agency for International Development (USAID) and managed by FHI 360, supports gender-sensitive programming, research, and learning to improve the economic security of highly vulnerable individuals, families, and children.

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 20 September 2018)

FHI 360 (2014) *Integration of Global Health and Other Development Sectors: A Review of the Evidence*, www.fhi360.org/resource/integration-global-health-and-other-development-sectors-review-evidence-summary-brief-full (accessed 13 September 2018)

Gupta, J. *et al.* (2013) 'Gender Norms and Economic Empowerment Intervention to Reduce Intimate Partner Violence Against Women

in Rural Côte d'Ivoire: A Randomized Controlled Pilot Study', *BMC International Health and Human Rights* 13.46: 1–12

Jewkes, R. *et al.* (2006) 'A Cluster Randomized-Controlled Trial to Determine the Effectiveness of Stepping Stones in Preventing HIV Infections and Promoting Safer Sexual Behaviour Amongst Youth in the Rural Eastern Cape, South Africa: Trial Design, Methods and Baseline Findings', *Tropical Medicine & International Health* 11.1: 3–16

Kim, J. *et al.* (2009) 'Assessing the Incremental Effects of Combining Economic and Health Interventions: The IMAGE Study in South Africa', *Bulletin of the World Health Organization* 87.11: 824–32

Kim, J.C. *et al.* (2007) 'Understanding the Impact of a Microfinance-Based Intervention on Women's Empowerment and the Reduction of Intimate Partner Violence in South Africa', *American Journal of Public Health* 97.10: 1794–802

Kirby, D.B.; Laris, B.A. and Rolleri, L.A. (2007) 'Sex and HIV Education Programs: Their Impact on Sexual Behaviors of Young People Throughout the World', *Journal of Adolescent Health* 40.3: 206–17

Leclerc-Madlala, S. (2008) 'Age-Disparate and Intergenerational Sex in Southern Africa: The Dynamics of Hypervulnerability', *AIDS* 22.suppl 4: S17–S25

Luke, N. (2005) 'Confronting the "Sugar Daddy" Stereotype: Age and Economic Asymmetries and Risky Sexual Behavior in Urban Kenya', *International Family Planning Perspectives* 31.1: 6–14

Pronyk, P.M. *et al.* (2006) 'Effect of a Structural Intervention for the Prevention of Intimate-Partner Violence and HIV in Rural South Africa: A Cluster Randomised Trial', *The Lancet* 368.9551: 1973–83

Swann, M. (2018) 'Economic Strengthening for HIV Prevention and Risk Reduction: A Review of the Evidence', *AIDS Care* 30.suppl 3: 37–84

UNICEF (2013) *Towards an AIDS-Free Generation – Children and AIDS: Sixth Stocktaking Report*, New York NY: United Nations Children's Fund

Wingood, G.M. *et al.* (2007) 'Efficacy of an HIV Prevention Program Among Female Adolescents Experiencing Gender-Based Violence', *American Journal of Public Health* 96.6: 1085–90

Wolbers, M. *et al.* (2011) 'Sample Size Requirements for Separating Out the Effects of Combination Treatments: Randomised Controlled Trials of Combination Therapy vs. Standard Treatment Compared to Factorial Designs for Patients with Tuberculous Meningitis', *Trials* 12: 26

This page is intentionally left blank

Glossary

3ie International Initiative for Impact Evaluation [India]
ACASI audio computer-assisted self-interview
ACET African Center for Economic Transformation [Ghana]
ASPIRES Accelerating Strategies for Practical Innovation and Research in Economic Strengthening
AVSI Association of Volunteers in International Service [Italy]
BRICS Brazil, Russia, India, China, and South Africa
CBA cost–benefit analysis
CCA cost–consequence analysis
CCF Child Community Family [Moldova]
CDD Community-Driven Development
CDI Centre for Development Impact [UK]
CEA cost-effectiveness analysis
CEDIL Centre of Excellence for Development Impact and Learning [UK]
CGD Center for Global Development [USA]
CIMMYT Centro Internacional de Mejoramiento de Maíz y Trigo [International Maize and Wheat Improvement Center, Mexico]
CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development, Brazil]
CUA cost–utility analysis
CV control village
DAC Development Assistance Committee
DAG Development Assistance Group
DALY disability-adjusted life year
DD/DiD difference-in-difference(s)
DFID Department for International Development [UK]
DRC Democratic Republic of the Congo
ECOSOC United Nations Economic and Social Council [USA]
ES economic strengthening
ESCAP Economic and Social Commission for Asia and the Pacific [Thailand]
ESFAM Economic Strengthening to Keep and Reintegrate Children in Families [Uganda]
EU European Union
FAPERJ Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro [Foundation for Research of the State of Rio de Janeiro, Brazil]
FaRe Family Resilience
GDP gross domestic product
GNI gross national income
GPEDC Global Partnership for Effective Development Cooperation
HHC Hope and Homes for Children [UK]
HLF High Level Forum
ICSU International Council for Science [France]
ICTD International Centre for Tax and Development [UK]
IDC international development cooperation

IDEAS International Development Evaluation Association [UK]
IDP integrated development programme
IFAD International Fund for Agricultural Development [Italy]
IFPRI International Food Policy Research Institute [USA]
IIED International Institute for Environment and Development [UK]
IR international relations
IRD integrated rural development
IRR internal rates of return
J-PAL Abdul Latif Jameel Poverty Action Lab [USA]
LE livre égyptienne [Egyptian Pound]
LEAP Livelihood Empowerment Against Poverty [Ghana]
LSHTM London School of Hygiene & Tropical Medicine
M&E monitoring and evaluation
MDG Millennium Development Goal
MeTA Medicines Transparency Alliance [The Netherlands]
MIT Massachusetts Institute of Technology [USA]
MQSUN Maximising the Quality of Scaling up Nutrition [UK]
MSC Most Significant Change
MV Millennium Village
MVP Millennium Villages Project
NBR net benefit ratio
NES National Electrification Scheme [Ghana]
NGO non-governmental organisation
NIH National Institutes of Health [USA]
ODA official development assistance
ODI Overseas Development Institute [UK]
OECD Organisation for Economic Co-operation and Development [France]
OED Operations Evaluation Department, World Bank [USA]
OPEC Organization of the Petroleum Exporting Countries [Austria]
PATH Program for Appropriate Technology in Health [USA]
PDIA problem-driven iterative adaptation
PEI Partnership for Economic Inclusion [USA]
PEPFAR President's Emergency Plan for AIDS Relief
PRA participatory rural appraisal
PRG peer review group
PSM propensity score matching
PUC-Rio Pontifical Catholic University of Rio de Janeiro [Brazil]
QALY Quality-adjusted life year
QCA qualitative comparative analysis
RCA Reality Check Approach
RCT randomised control trial
RISE Research on Improving Systems of Education [UK]
RMNCH reproductive, maternal, newborn, and child health
SADA Savannah Accelerated Development Authority [Ghana]
SDG Sustainable Development Goal
SSC South–South cooperation
STI sexually transmitted infection
SWPRP Southwest Poverty Reduction Project [USA]

TBIE theory-based impact evaluation
TCDC Technical Cooperation among Developing Countries
ToC theory of change
ToR Terms of Reference
TOSSD Total Official Support for Sustainable Development
TZ *tuo zaafi* [millet staple dish, Ghana]
UNCTAD United Nations Conference on Trade and Development [Switzerland]
UNDCF United Nations Development Cooperation Forum [USA]
UN DESA United Nations Department of Economic and Social Affairs [USA]
UNDP United Nations Development Programme [USA]
UPS upgrading strategies
USAID United States Agency for International Development
VFM value for money
VSLA village saving and lending association
WFP World Food Programme [Italy]
WHO World Health Organization [Switzerland]

This page is intentionally left blank

Institute of Development Studies

Volume 49 (2018)
No. 1 Value Chains for Nutrition in South Asia: Who Delivers, How, and to Whom?
No.1A Inclusive Peace and Security
No. 2 Accountability for Health Equity: Galvanising a Movement for Universal Health Coverage
No. 3 Emerging Economies and the Changing Dynamics of Development Cooperation

Volume 48 (2017)
No. 1 Sex Education in the Digital Era
No. 2 Interrogating Decentralisation in Africa
No. 3 Africa's Youth Employment Challenge: New Perspectives
No. 4 Courting Catastrophe? Humanitarian Policy and Practice in a Changing Climate
No. 1A Has Universal Development Come of Age?
No. 5–6 Green Power for Africa: Overcoming the Main Constraints

Volume 47 (2016)
No. 1 Opening Governance
No. 1A Connecting Perspectives on Women's Empowerment
No. 2 Development Studies – Past, Present and Future
No. 3 Ruptures and Ripple Effects in the Middle East and Beyond
No. 4 Foresight in International Development
No. 5 Power, Poverty and Inequality
No. 2A States, Markets and Society – New Relationships for a New Development Era
No. 6 Engaged Excellence

Periodical ID statement The *IDS Bulletin* (ISSN 0265-5012 print) is published bi-monthly in January, March, May, July, September and November. Mailing agent: Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK.
Tel: +44 (0)1273 606261 idsbulletin@ids.ac.uk

Postmaster Send all address changes to *IDS Bulletin*, Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK.
Tel: +44 (0)1273 606261 idsbulletin@ids.ac.uk

Journal customer services For ordering information, claims and any enquiry concerning your journal subscription please contact Gary Edwards, Institute of Development Studies, Library Road, Brighton BN1 9RE, UK. Tel: +44 (0)1273 606261 idsbulletin@ids.ac.uk

Print details Printed in the UK by Nexus Design & Print Ltd, Studio 38, Adur Business Centre, Little High Street, Shoreham-by-Sea BN43 5EG, UK.

Delivery terms and legal title Prices include delivery of print journals to the recipient's address. Delivery terms are Delivery Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch.

ISSN 0265-5012 (print), 1759-5436 (online)

IDSBulletin

Transforming Development Knowledge

Institute of Development Studies

The Millennium Villages: Lessons on Evaluating Integrated Rural Development

Editor **Chris Barnett**

Volume 49 | Number 4 | September 2018

IDS Bulletin
Transforming Development Knowledge

Foreword
Richard Longhurst

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

Integrated Development, Past and Present
Edoardo Masset

The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies
Edoardo Masset, Giulia Mascagni, Arnab Acharya, Eva-Maria Egger and Amrita Saha

Assessing Value for Money in Integrated Development Programmes – The Case of a Millennium Villages Project in Northern Ghana
Arnab Acharya and Tom Hilton

Abductive Reasoning to Explain Integrated Development: Lessons from the Multi-Method Evaluation of the Millennium Villages Project
Dee Jupp and Chris Barnett

Can Immersion Research Add Value in Understanding Integrated Programme Interventions?
Dee Jupp, David Korboe and Tony Dogbe

Learning About Integrated Development Using Longitudinal Mixed Methods Programme Evaluation
Emily Namey, Lisa C. Laumann and Annette N. Brown

Applying Factorial Designs to Disentangle the Effects of Integrated Development
Holly M. Burke, Mario Chen and Annette N. Brown

‘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.’

ISSN 0265-5012 (print), 1759-5436 (online) DOI: 10.19088/1968-2018.155