

Getting the most out of impact evaluation for learning, reporting and influence

Insights from piloting a Participatory Impact Assessment and Learning Approach (PIALA) with IFAD

by
Edward Heinemann
IFAD

Adinda Van Hemelrijck
Institute of Development Studies, University of Sussex

Irene Guijt
Oxfam GB

16 IFAD
RESEARCH
SERIES



The IFAD Research Series has been initiated by the Strategy and Knowledge Department in order to bring together cutting-edge thinking and research on smallholder agriculture, rural development and related themes. As a global organization with an exclusive mandate to promote rural smallholder development, IFAD seeks to present diverse viewpoints from across the development arena in order to stimulate knowledge exchange, innovation, and commitment to investing in rural people.

The opinions expressed in this publication are those of the authors and do not necessarily represent those of the International Fund for Agricultural Development (IFAD). The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of IFAD concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The designations “developed” and “developing” countries are intended for statistical convenience and do not necessarily express a judgement about the stage reached in the development process by a particular country or area.

This publication or any part thereof may be reproduced for non-commercial purposes without prior permission from IFAD, provided that the publication or extract therefrom reproduced is attributed to IFAD and the title of this publication is stated in any publication and that a copy thereof is sent to IFAD.

Authors:

Edward Heinemann, Adinda Van Hemelrijck, Irene Guijt

© IFAD 2017

All rights reserved

ISBN 978-92-9072-767-5

Printed July 2017



Getting the most out of impact evaluation for learning, reporting and influence

Insights from piloting a Participatory Impact Assessment and Learning Approach (PIALA) with IFAD

by
Edward Heinemann
IFAD

Adinda Van Hemelrijck
Institute of Development Studies, University of Sussex

Irene Guijt
Oxfam GB



16 IFAD
RESEARCH
SERIES

Acknowledgements

The PIALA initiative was made possible with financing from the DFID-funded Innovation Mainstreaming Initiative and the Bill & Melinda Gates Foundation. It was initiated and headed by one of the authors, Edward Heinemann (then Senior Strategic Planning Officer, IFAD). The methodological design and piloting was led by the second author, Adinda Van Hemelrijck (consultant for IFAD), with guidance and support provided by Irene Guijt, the third author, plus two other consultants, Jeremy Holland and Andre Proctor. During the initiative, valuable feedback was received from IFAD colleagues and from four independent global experts (Carlos Barahona, Robert Chambers, Marie Gaarder and Kent Glenzer). The authors are also grateful for the extensive and helpful comments provided by Ulaş Demirag (then Country Programme Officer for IFAD in Ghana) on an earlier draft of this paper.

About the authors

Edward Heinemann currently works as a lead technical specialist in the Policy and Technical Advisory Division, promoting IFAD's country-level policy engagement agenda. He was the principal author of the 2014 Consultation Report for the Tenth Replenishment of IFAD's resources, and he was previously the team leader for, and one of the main authors of, IFAD's 2011 *Rural Poverty Report*. Before this he worked in IFAD's Change and Reform unit, where he authored the Strategic Framework 2007-2010, IFAD's policy on sector-wide approaches, its partnership strategy and its grants policy. Prior to 2007, Edward worked in operations, initially as a country programme manager for several countries in sub-Saharan Africa, and later as the Regional Economist for East and Southern Africa. Before joining IFAD, he worked at the African Development Bank as a project officer. He holds an MA in rural development from the University of East Anglia in the UK.

Adinda Van Hemelrijck is an independent consultant working on design, evaluation and learning around impact in collaborative settings in international (mostly rural) development. Her work focuses on strengthening local capacity and enabling multi-stakeholder engagement in impact inquiry to foster social innovation and transformative change. Her recent work has been with IFAD, UNICEF, WHO, Oxfam GB, the International Institute for Environment and Development, Oxfam GB and the Catholic Relief Service. From 2007 to 2012, she served as a global senior M&E advisor at Oxfam America to assist regional offices and partners in developing collective impact evaluation and learning frameworks taking a rights-based approach. In the 10 years prior to this, she worked with various EU-based agencies – including the Belgian Special Evaluation Office – on performance evaluation and learning, gender and communication for social change. Adinda holds an MA in comparative cultural science and international relations from the University of Gent in Belgium, and is pursuing a doctoral degree at the Institute of Development Studies, University of Sussex, UK.

Irene Guijt leads Oxfam GB's Research Team, which generates evidence to influence economic, environmental and social justice, and improves the overall quality of research at Oxfam. Prior to joining Oxfam in 2015, Irene worked for 25 years in rural development, natural resource management, collective action and social justice – and particularly in making less heard voices more audible and influential. Her recent work includes pioneering the SenseMaker® stories-at-scale approach in international development for impact evaluation; building global evaluation capacity through BetterEvaluation; working on theory of change for transformational development with Hivos; and co-convening the Big Push Forward, including co-editing the book *The Politics of Evidence and Results*. She has also co-hosted the annual “M&E on the Cutting Edge” series for six years. Irene previously worked at the International Institute for Environment and Development, was a Visiting Fellow at the Australian National University and a Research Associate for the Overseas Development Institute. Irene holds a BSc and an MSc in land and water use engineering from Wageningen University and Research Centre, the Netherlands. Her PhD focused on small-scale producers in Brazil, and how messy partnerships among civil society, government and research can lead to learning and innovation in food systems.

Oversight: Paul Winters, Director of Research and Impact Assessment Division, and Ashwani Muthoo, Director of Global Engagement, Knowledge and Strategy Division.

Advisory Board: Fabrizio Bresciani, Shirley Chinien, Edward Heinemann, Bruce Murphy, Richard Pelrine, Lauren Phillips, Tomas Rosada and Abdelkarim Sma.

Editorial Management Team: Rui Benfica, Helen Gillman and Anja Lesa.

Table of contents

Acknowledgements	2
Abstract	5
Scanning for options for impact evaluation	6
PIALA principles, elements and standards	8
What emerged from the evaluations that piloted PIALA	15
Reflections on the approach in evaluations that piloted PIALA	21
First phase – focusing and framing the evaluation	21
Second phase – collecting and linking data.....	23
Third phase – analysing contributions	27
Lessons learned and critical next steps for PIALA	30
References	33

Abstract

This paper reflects on the use and usefulness of a systemic theory-based and participatory mixed-methods approach for evaluating the complex contributions of IFAD-funded projects to rural livelihoods and poverty in ways that involve stakeholders in reflection and learning. The Participatory Impact Assessment and Learning Approach (PIALA) was developed and piloted under IFAD's Innovation Mainstreaming Initiative with additional funding from the Bill & Melinda Gates Foundation. The paper describes the specific features of the approach, illustrated with cases from Viet Nam and Ghana, in order to help commissioners of evaluations assess its potential relevance. The paper assesses the value of the approach for collaborative learning and reporting for IFAD's country programming and global policy engagement, as well as for the wider development community.

Scanning for options for impact evaluation

Over the past 15 years or so, the international development community has increasingly sought to improve development effectiveness. The growing desire to demonstrate and measure impact has created a demand for, and stimulated the production of, impact evaluation.

IFAD, too, has sought to measure and learn about the impact of its investments on rural poverty. In 2012, it committed both to lifting 80 million rural people out of poverty during its Ninth Replenishment period (2013-2015) and to measuring its performance in working towards this end. However, despite improvements in the quality of monitoring and evaluation (M&E) in IFAD-supported development projects, and the work of IFAD's Independent Office of Evaluation, and the parallel internal self-evaluation system, getting reliable data, quantitative and qualitative, on outcomes and impacts for most projects remained a challenge. Therefore, IFAD made a further commitment to conduct approximately 30 impact assessments, several of which would be randomized control trials (RCTs) or quasi-experimental impact studies.

The organization was also interested in methodological experimentation. A number of related factors drove this interest. First, there was some uncertainty as to the value and replicability of (quasi-) experimental studies for assessing the impact of complex processes of rural poverty reduction (see Box 1). Second, it was recognized that there was a need to expand and deepen IFAD's knowledge of what works for poor rural people, under which conditions, for whom, why and how – particularly how different efforts interact and combine to generate development outcomes and impacts – or not. Third, there was interest in promoting increased levels of collaboration between IFAD, governments and other public and private actors involved in project implementation and management – and particularly in learning about the changes that the projects induce in smallholder production and market systems and the impact of these changes on rural poverty (IFAD, 2011, 2012). Fourth, there was a concern to extend IFAD's people-centred approach (Guijt and Woodhill, 2002), characterized by participatory approaches in project design and implementation, into the realm of more participatory approaches for impact assessment, while at the same time avoiding compromising on rigorous causal analysis (Chambers, 2015).

In essence, the challenge was to develop a feasible and replicable approach for impact evaluation that could produce rigorous evidence – both qualitative *and* quantitative – and at the same time support collaborative learning and promote improved understanding of the processes and pathways of socio-economic change. The expectation was that such impact evaluation could contribute to improving: (a) country programme and project design and implementation; (b) policy engagement around agricultural and rural development at both the global and the country levels; and (c) measurement and reporting of results and development impact (IFAD, 2011).

Box 1: Options for impact evaluation

Traditional counterfactual-based approaches using (quasi)experimental methods are very effective in rigorously attributing the effect of a specific mechanism or subcomponent of a project. But they are generally less suitable for more complex environments with high causal density, spillover, time lags and emergent causal paths (Picciotto, 2012; Woolcock, 2009). They focus too narrowly on specific intervention components, thus “leaving many evaluation questions unanswered” (White, 2014, p. 3; Winters and Garbero, 2015). They do not explain impacts or investigate sustainability, given their focus on isolated cause-effects, and therefore they do not indicate whether, how or why similar relations could possibly work out elsewhere (Woolcock, 2013). Finally, engagement and learning with partners and stakeholders is inhibited by scientific procedures (Rogers and Peersman, 2014).

Alternative theory-based and complex systems approaches, on the other hand, are generally quite time-intensive and produce evidence that cannot be compared across many cases. They are therefore not suitable for larger populations (Beach and Pedersen, 2013). In addition, those that allow for participation generally do not set out to rigorously assess causality and address concerns of bias (Copestake, 2013; White and Phillips, 2012).

Source: Van Hemelrijck, 2016b.

To address these challenges, in 2012 IFAD launched an innovation initiative for developing a Participatory Impact Assessment and Learning Approach (PIALA).¹ The initiative had three objectives. The first was to design and test a cost-effective, potentially scalable participatory approach for impact evaluation and learning that could produce rigorous qualitative and quantitative evidence, complement existing approaches and strengthen IFAD’s self-evaluation system. The second was to conduct two pilot impact evaluations using this approach to generating substantive understanding of the nature and extent of the impact of IFAD investments. The two cases selected for this were the Developing Business with the Rural Poor (DBRP) programme in Ben Tre province, southern Viet Nam, and the national Root and Tuber Improvement and Marketing Programme (RTIMP) in Ghana. The third objective was to facilitate project-level and global stakeholder reflections on the pilots. The lessons learned could then be shared to generate more interest in the approach both within IFAD and among other international development agencies and the wider global development community (IFAD and BMGF, 2013b).

The fields of social research and impact evaluation offer an abundance of tested designs and methodologies. The intent therefore was not to reinvent the wheel but to think through how existing designs and methodologies could be combined in novel and creative ways to meet IFAD’s needs. Inspiration was drawn mostly from the theory-based (in particular, realist) and transformative (including rights-based) traditions (Chambers, 2008; Holland, 2013; Mertens, 2009; Pawson, 2013; Van Hemelrijck, 2013).

1. All publications relating to the PIALA initiative can be found on the IFAD website: www.ifad.org/topic/overview/tags/piala.

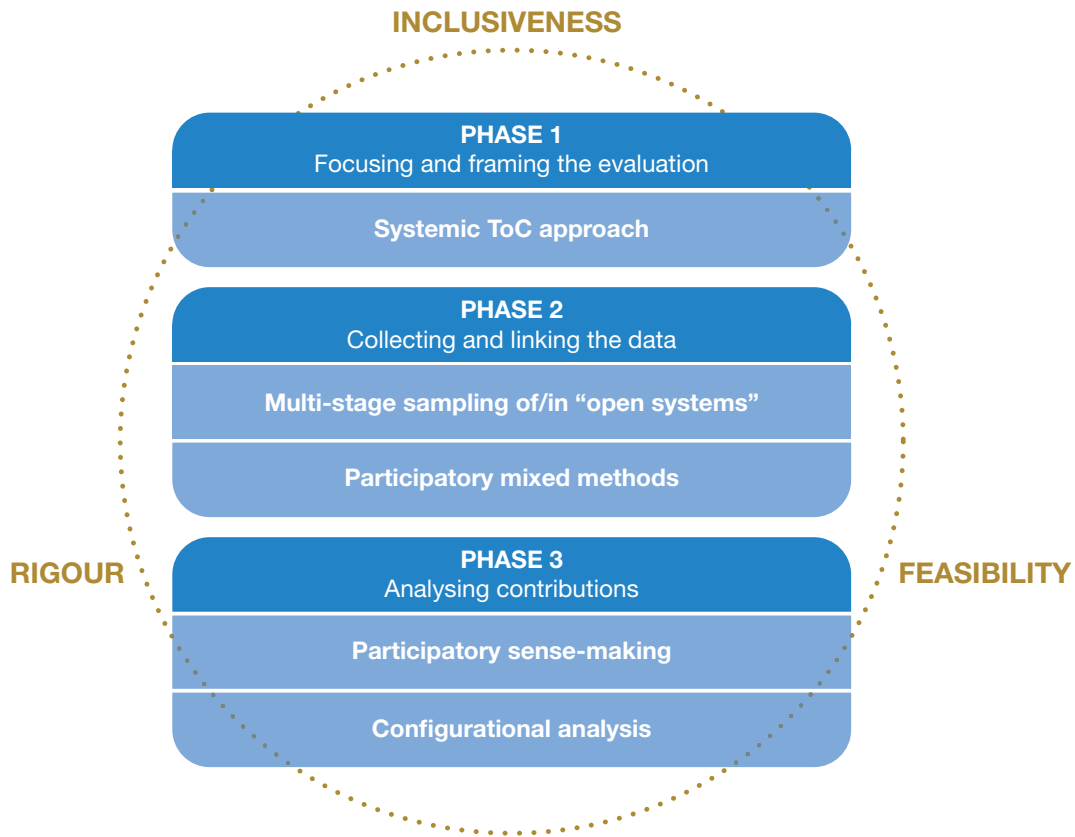
PIALA principles, elements and standards

PIALA has been developed as an approach for assessing, explaining and debating – in a participatory and rigorous way – contributions to impact of medium-n to large-n projects or programmes. It can be used in contexts where conventional counterfactual-based approaches do not apply or are difficult to pursue. It has three purposes: *reporting* to what extent impact occurred or not; stimulating collaborative *learning* among different stakeholders around why impact did or did not occur; and so *influencing* policies and strategies for rural poverty eradication to make them more effective, inclusive and sustainable. To serve these three purposes, PIALA needs to produce conclusions, based on rigorous evidence, that can be generalized for the entire impacted population. Different stakeholders must be engaged in building and analysing the evidence base, and in debates organized around the emerging evidence.

For this PIALA draws on two *design principles*: (a) think and evaluate systemically; and (b) enable meaningful participation in building, debating and analysing evidence. These two principles are essential for making an impact evaluation more useful in terms of its learning value, policy influence and accountability. An evaluation using PIALA can adopt any mix of designs and methods as long as it remains consistent with these two basic principles. Or as Stern puts it (2015, p.13), “in [impact evaluation] in particular, it may be mixed designs rather than mixed methods that are most useful. Often what are required are several well-chosen designs, each of which will use a variety of methods and be tailored to answer the various [impact evaluation] questions posed by the commissioners and other stakeholders. Few evaluations ask a single question.”

PIALA views impact from a systemic perspective, as the system of interactions between various actors and influences, rather than the direct and isolated relationship between intervention and effect. A systemic view seeks to move beyond assessing “what has changed” to also answer the more difficult “how” and “why” questions and evaluate the likely sustainability of the observed change (Befani et al., 2015; Burns, 2014). PIALA thus typically addresses the questions: what has changed, for whom and how; how sustainable are the changes likely to be; what are the impacts of these changes, and what caused them; what contributions made by the project/programme are being evaluated; and what are the implications for policy, strategy and targeting (IFAD and BMGF, 2013b).

Figure 1: PIALA elements and standards



If budget and capacity limitations make it difficult to investigate these questions, and serve all three purposes exhaustively, then choices have to be made. Which questions and purposes matter most and deserve greatest attention depends on the evaluation context and requirements (Rogers, 2009; Stern, 2015). The PIALA pilots have shown that trade-offs can often be turned into win-wins, even with limited resources, by spending enough time up front with the sponsors to carefully consider the different design options and explore how these could generate greater value (Van Hemelrijck and Guijt, 2016).

PIALA has a *quality framework* that combines standards of *rigour*, *inclusiveness* and *feasibility* to guide these design decisions and assess the value of an evaluation in the three main evaluation phases – focusing and framing the evaluation; collecting and linking the data; and analysing contributions to impact. *Rigour* is understood as being the quality of thought put into the methodological design and conduct of the evaluation, beyond any particular procedure, in

order to overcome bias and ensure consistency and responsiveness (Rogers, 2009; Stern et al., 2012). It goes beyond mainstream evaluation practice, where rigour denotes the controlled avoidance of bias through statistical procedure (Bamberger and White, 2007); and thus is insufficient for mixed-methods designs using different methodologies. PIALA draws on the premise that bias cannot be avoided by using a single method or procedure, but must be mitigated through triangulation of different methods and perspectives (Befani et al., 2014; Camfield et al., 2014; Mertens, 2009). *Inclusiveness* refers to the legitimacy of the ways in which people are engaged in the evaluation, and to the level of impartiality or inclusion of all stakeholder views and perspectives. It has an intrinsic empowerment value but also contributes to the robustness and credibility of the evidence, and thus to the validity of the findings (Chambers, 2015; Pawson, 2013). *Feasibility* concerns the budgetary resources, skills and capacity needed to meet the expectations of rigour and inclusiveness and enhance learning (IFAD and BMGF, 2015).

As shown in Figure 1, PIALA consists of **five key elements** that combine a variety of tested methodologies in the three main phases of evaluation (i.e. focusing and framing, collecting and linking data, and analysing contributions) to assess systemic impact on a larger scale:

1. A *systemic Theory of Change* approach to help visualize the causal claims and enable stakeholders to engage in the framing of the evaluation and debating the evidence.
2. *Multi-stage sampling of/in "open systems"* to enable systemic inquiry across medium-to-large populations.
3. A semi-standard set of *participatory mixed methods* to collect and link quantitative and qualitative data in the sampled "systems" in a systematic and comparable way.
4. A *participatory sense-making* model that engages stakeholders at the local and aggregated levels in debating emerging evidence.
5. A *configurational analysis* method to assess systemic change patterns and causality, and draw conclusions about the distribution and magnitude of impact, across medium-to-large populations.

The first of the PIALA elements – the *systemic Theory of Change (ToC)* approach – forms the backbone of the entire evaluation. It is especially useful for evaluating systemic change that involves many different interventions, implementers, contributors and funders, as it helps create a shared understanding of complex pathways and enables stakeholders to critically engage in the analysis. It involves reconstructing and visualising project impact pathways and change hypotheses, together with the broader trends and influences, based on a thorough desk review and discussions with key stakeholders. The desk review typically looks at the project design paper, logical framework, M&E and supervision reports, and other relevant studies. Discussions are held to establish what potential users of the evaluation already know, and what they expect to see coming out of the evaluation. This helps ensure that the evaluation responds to their knowledge needs.

On the basis of the desk review and discussions, a causal flow diagram is produced showing the systemic links and feedback loops between the subsystems the project sought to influence; the mechanisms it triggered or developed for this; the collective outcomes of these and other influences; and the resulting impact on people's lives (ibid.). The diagram is constructed by travelling from the impact back through the presumed causal pathways to the project mechanisms. It builds on project theory (e.g. the logical framework) but looks in the opposite direction and takes a broader view of impact that includes non-project influences. The project theory, by contrast, is constructed from a management perspective. It thus focuses on strategy and performance that look towards delivering the planned results that are expected to lead to the desired impact (Funnell and Rogers, 2011).

The second PIALA element – *multi-stage cluster sampling of/in "open systems"* – ensures that the various populations are sufficiently representative to enable comparison and generalization of findings. Cluster sampling is generally more cost-effective than other random sampling strategies. It substantially reduces costs and logistics, and makes it possible to nest different data collection methods suitable for evaluating complex projects (IFAD and BMGF, 2013b; Levy and Barahona, 2002). At IFAD, the system that forms the principal unit of analysis, and thus the first population to sample, usually concerns agricultural production and market systems developed at a medium-to-large scale. From the principal sample of these systems, subsamples are drawn representing the different populations affected by the various parts of the systems. The populations are stratified and mostly sampled randomly (Van Hemelrijck, 2016b). For quantitative poverty analysis, the total size of household subsamples must be large enough for statistical analysis (with at least a 95 per cent confidence interval). For the comparative analysis of systemic interactions affecting poverty, the subsamples of intended beneficiaries participating in focus group discussions must together sufficiently cover the main variations in conditions and treatments under the different project mechanisms across the entire project area.

The third PIALA element – *participatory mixed methods* – involves the appropriate selection and mixing of methods and processes, mostly participatory, to collect qualitative and quantitative data on the different causal links in the ToC. A core set of methods has been developed and tested in the IFAD pilots for agricultural value chain and market systems projects. The selected methods should make it possible to build the actual causal paths with evidence that validates (or refutes) the ToC, and compare the causal paths across the entire sample, as the basis for aggregating findings for the project as a whole. The participatory methods therefore need to be more or less standardized for the entire sample, yet sufficiently open-ended to draw out various causal and contextual elements. The methods should also complement and build on each other analytically, each focusing on different causal links in the ToC to investigate the cascading causes and effects in the ToC and build the causal paths. They also need to partly overlap to permit triangulation, thus allowing inquiry one step further up or down the causal chain. Data collation alongside the ToC makes it possible to cross-check and link data for building the causal paths during fieldwork and identify data gaps or weaknesses needing further inquiry in a timely manner (IFAD and BMGF, 2015).

Table 1: Complementary and overlapping methods

Level of change				
	Reach and effects of project mechanisms (DSF, FFF, GPC, MEF) $M_i \rightarrow C_i$	Roots and tubers livelihood changes and causes $C_i \rightarrow O_i$	Household poverty (foods, assets and income) $\Sigma O_i \rightarrow I2 \rightarrow I1$	
Methods			Household survey (n 840) on changes in food, assets, income and roots and tubers revenues	
		Generic change analysis (n 439) (ranking and causal flow mapping of changes in wealth and wellbeing)		
		Livelihood analysis change (n 400) (matrix scoring and causal flow mapping)		
		SenseMaker® (n 393) (self-signification of micro-narratives)		
		Constituent feedback (n 341) With DSF, FFF and GPC/MEF beneficiaries		
		Key informant interviews (n 100) With officials and service providers		

Table 1² illustrates how this mixed-methods logic worked out in the IFAD pilot in Ghana. The codes in the top row reflect the types of causal link, which in the actual RTIMP ToC were much more complex (with the numbering reflecting the contribution claim and level in the ToC).³ A classic household survey and a participatory method for generic change analysis (right side of the table) were used – the latter in gender-specific focus groups – to take stock of the impacts of changes in livelihoods on household food and income.⁴ A livelihood change analysis method, used in different but also gender-specific focus groups, was triangulated with the generic change analysis (middle of the table) to further investigate the effects of the observed livelihoods changes and the changes in production

2. Source: Presentation of results and reflections from the impact evaluation of the RTIMP evaluation (IFAD, Rome, 26 October 2015). Available at: www.ifad.org/documents/10180/d791c8af-51f2-4768-afef-49caf12b7ae8. More information on the specific hypotheses, links and questions examined by each of these methods can be found in the evaluation design paper, available at: www.participatorymethods.org/sites/participatorymethods.org/files/IFAD-RTIMP%20Impact%20Evaluation%20Design%20Paper%20Dec%202014.pdf.

3. The “M” in the codes stands for “mechanism” and refers to the institutional or organizational arrangements developed by the project to bring about the desired changes in capacities and behaviours of the supply chain actors. These mechanisms are the district stakeholder forums (DSFs), farmer field forums (FFFs), good practice centres (GPCs) and Micro-Enterprise Fund (MEF). The “C” in the codes refers to these changes, while the “O” indicates the outcomes of these changes leading to the ultimate impacts on rural households indicated by the “I”. The numbering in the codes reflects the contribution claim and level in the ToC.

4. Poverty impact is measured by IFAD in terms of household assets, income, food and children’s nutrition (IFAD, 2005, 2011d). The impact survey questions in IFAD’s Results and Impact Management System (RIMS) related to diet diversity or daily intake requirements, and children’s malnutrition or anthropometrics, have been left out of the PIALA pilots. These questions themselves require a statistical study that is quite resource-intensive and was not considered useful by the commissioners in the context of the DBRP and RTIMP impact evaluations.

Box 2: Participatory rural appraisal and SenseMaker®

Participatory rural appraisal refers to a set of participatory research methods that were developed and became widely used in the 1990s. In the IFAD pilot in Ghana, a limited SenseMaker® study was also conducted to capture unintended effects and influences, and uncover the deeper motivational dynamics affecting relations and changes in livelihoods. SenseMaker® is a methodology with patented software developed by Cognitive Edge. It facilitates mass ethnography and provides a way of nearly real-time mapping of social interactions and individual perceptions and motivations to inform adaptive management and policy formulation. It collects large amounts of self-signified micro-narratives or fragmented stories capturing people's experiences and how these shape their perceptions of past and future change in ways that enable us to identify emerging patterns of actions and decisions. The software also permits statistical analysis on a very large scale. While its strength lies in real-time monitoring of emerging impact, it was used in the PIALA pilot in Ghana to test its value for ex post impact evaluation.

For more information about SenseMaker®, see:
www.sensemaker-suite.com

For more information about participatory rural appraisal, see:
https://en.wikipedia.org/wiki/Participatory_rural_appraisal

and market systems that caused them. In contexts where the “system” corresponds to the local level of governance, such as a village or local farmer or membership organization, a relationships analysis can be conducted (again, with different but gender-specific groups) to investigate access to institutions, services and supplies. This was done in the PIALA evaluation in Viet Nam (IFAD and BMGF, 2014), as well as a subsequent PIALA evaluation conducted in Myanmar (Van Hemelrijck, 2017b). The tools used in these two methods were largely based on participatory rural appraisal (see Box 2). Constituent feedback (see Box 3) was further triangulated with livelihood analysis (left side of the table) to investigate the functioning, reach and effects of project mechanisms on the changes in the various areas affecting livelihoods. Finally, key informant interviews were conducted with local, regional and national project stakeholders to cross-check the evidence from the constituent feedback and livelihood analysis.

The fourth PIALA element – *participatory sense-making* – involves small local workshops with 20 to 30 participants organized during the fieldwork at each of the sites, and a project-level workshop with over 100 participants that takes place after fieldwork has been completed but before the final analysis. At these sense-making workshops, stakeholders discuss the evidence and assign value to project contributions (among other influences). They do this by comparing the actual causal paths revealed by the evidence with those hypothesized in the ToC. Participants are selected to include all the different perspectives necessary to cross-validate the evidence and inform the final analysis (e.g. decision makers, service providers, intended beneficiaries and other relevant actors). Sense-making at these two levels helps to cross-check and strengthen the evidence, gives participants a voice and ownership, and encourages stakeholders' systemic learning and collective responsibility. In essence, it makes an evaluation more democratic: even small moments of participation can add up to ripples of empowerment (Van Hemelrijck, 2016c).

Box 3: Constituent feedback

Constituent feedback (also known as constituent voice) is a methodology developed by Keystone Accountability (www.keystoneaccountability.org). It is used for collecting quantified feedback and engaging in dialogue with key constituents or beneficiaries using standardized metrics similar to customer satisfaction surveys developed in the private sector and descriptive statistics to produce visual data reports. While its strength lies in participatory performance measurement during project implementation, it was used in the PIALA pilot in Ghana to test its value for ex post impact evaluation in assessing and critically debating the reach of programme mechanisms and their benefits and contributions to livelihood changes. Similar to SenseMaker®, constituent feedback also collects and analyses large amounts of self-scored feedback data within a very short space of time and at an even lower cost. It focuses on the performance of specific service delivery mechanisms, however, unlike SenseMaker®, which investigates broader patterns and trends of change.

The constituent feedback method can be replaced by other organizational assessment methods. In a subsequent PIALA evaluation of an Oxfam GB programme on local governance for resilience in Myanmar, a performance score card was used instead of constituent feedback to assess the level of sustained effectiveness of community membership organizations 18-20 months after the project had closed. The scorecard had been developed and used for monitoring, and was adapted to fit the evaluation (Van Hemelrijck, 2016a).

The evaluation report is available at:

<http://policy-practice.oxfam.org.uk/publications/governance-in-myanmar-evaluation-of-the-building-equitable-and-resilient-leveli-620177>

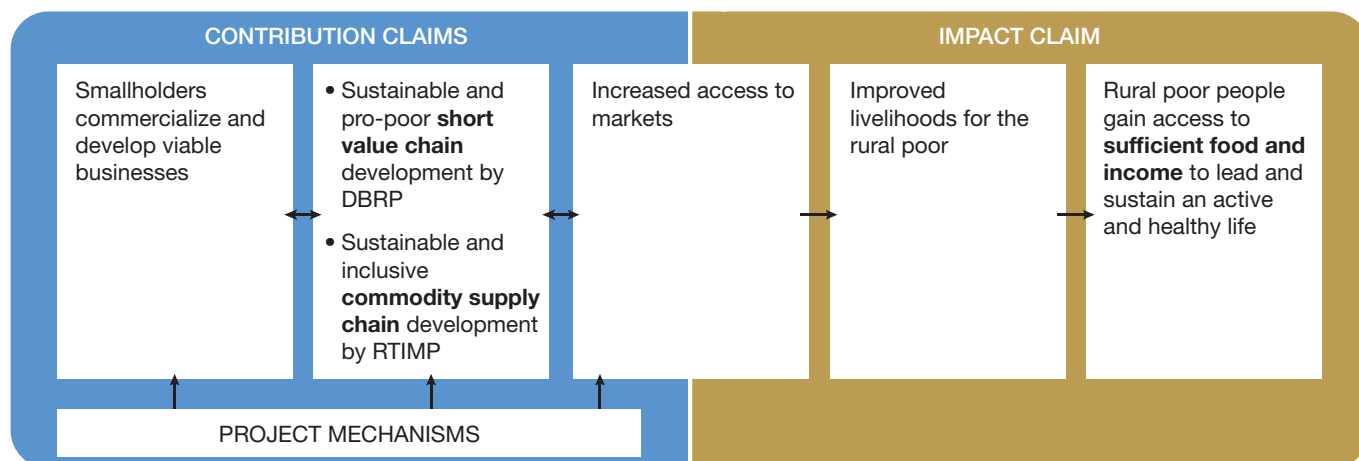
The last PIALA element – *configurational analysis* – involves comparing systemic change and impact across sample systems to reach conclusions about the distribution and magnitude of impact. There are four basic steps in the analysis: (1) aggregated collation of all field collation matrices and relevant data from secondary sources, including project M&E; (2) clustering and comparison of evidence to identify patterns or configurations of system change and causal attributes; (3) comparative analysis of similarities and differences in configurations for specific mechanisms or parts of the system (including cases with or without functioning mechanisms; and (4) zipping up the findings against the ToC to draw conclusions about the distribution and magnitude of how the project contributed to the impact. The analysis uses a combination of contribution scoring, process tracing and cross-tabulation techniques (IFAD and BMGF, 2015; Van Hemelrijck, 2016c). Software such as EvalC3⁵ can be used to assess the conjuncture of different mechanisms and causal processes, and identify those that perform best (Van Hemelrijck, 2016a).

5. This is a novel software (<https://evalc3.net/>), which was piloted in an impact evaluation using PIALA for Oxfam GB in Myanmar. The software helps to identify sets of causal attributes revealed by the evidence across the sample that are *necessary* and/or *sufficient* for specific sets of outcome attributes to occur. It also compares and evaluates the performance of these causal models in order to find those that have the greatest predictive power.

What emerged from the evaluations that piloted PIALA

The two projects evaluated. PIALA was piloted in the Developing Business with the Rural Poor Programme (DBRP) in Viet Nam,⁶ and the Root and Tuber Improvement and Marketing Programme (RTIMP) in Ghana.⁷ These two projects were selected because of their focus on market and value chain development; their level of maturity (five years of implementation); their level of disbursement of IFAD funding (over 70 per cent); and their track record and what could be learned from them about scaling up across different contexts (IFAD and BMGF, 2013b). While DBRP focused on developing diversified short value chain systems, RTIMP sought to develop longer commodity chains linked to national and export markets and industries. The RTIMP evaluation focused on four main commodities: *gari*, high-quality cassava flour, plywood cassava flour and fresh yam for export. Both projects aimed at improving livelihoods and increasing food and income security by enhancing smallholders' capacity to commercialize these commodities and linking local businesses to markets and industries. Figure 2 shows the main hypothesis at the heart of the ToC of both projects and Box 4 provides a narrative description (IFAD and BMGF, 2014; MOFA/GOG et al., 2015).

Figure 2: Common hypothesis of DBRP and RTIMP



6. The DBRP was implemented from 2008 to 2014 in two provinces (Cao Bằng and Bến Tre) with a total budget of US\$51 million, including a US\$36 million loan from IFAD. The evaluation was conducted in 2013 at a cost of US\$90,000 and only in Bến Tre province, where the project was implemented in 50 of 164 communes in eight of nine districts (IFAD and BMGF, 2014). The evaluation report is published on IFAD's website: www.ifad.org/documents/10180/841677fb-7329-462e-866d-efbaf5edc6cb.

7. The RTIMP was implemented from 2007 until 2015 as a national programme in 106 of 216 districts spread across all ten regions countrywide, with a total budget of US\$24 million, of which US\$19 million was financed under an IFAD loan. The evaluation was conducted countrywide after project completion in 2015, at a cost of US\$233,000, and covered the post mid-term review period from 2010 (MOFA/GOG et al, 2015). The evaluation report is published on IFAD's website: www.ifad.org/documents/10180/7b74a2e6-e4bc-4514-a99e-44e0ee9adb7f.

Box 4: Theory of Change (ToC) narratives for the Developing Business with the Rural Poor Programme (DBRP) and the Root and Tuber Improvement and Marketing Programme (RTIMP)

Under DBRP, locally viable and pro-poor short value chains (such as ornamental trees, decorative plants, fruit, livestock, shrimps and coconut) were expected to provide an opportunity for developing small businesses. In turn, this was expected to generate new livelihood opportunities for resource-poor households. The assumption was that local businesses would be able to take advantage of an improved entrepreneurial and business environment, and that they would be willing and able to work with poor households, offering them jobs and sharing information on how to operate in the market. DBRP pursued three important contribution claims to achieve this – with common interest groups (CIGs), socio-economic development planning (SEDP), commune investment funds (CIFs), public-private partnership models (PPP) and government-led market and business support services as the main mechanisms that would trigger the desired changes of these claims:

- *Enhanced market access at the village and the commune level.* The development of CIGs would enable small farmers and business households to market their products, while participatory SEDP would empower them to influence local government decisions and allocations. The CIF would help create greater access to markets by providing funding for local infrastructure development, training and services offered by local organizations such as women's and farmers' associations.
- *Enhanced market and job creation at the district and the provincial level.* Strengthening the capacity of local government units to provide business support services and developing provincial facilities to support small and medium-sized businesses in the rural districts would help establish the enabling environment needed to boost the market and business opportunities and create new jobs.
- *PPPs at the provincial level.* A special credit model to fund short value chains would help everyone in the chain access better business finance and support services and thus mitigate risks. This would stimulate greater investment in new smallholder businesses in value chains and help small and medium-sized businesses in the districts grow.

The RTIMP assumed that livelihoods and poverty status could be improved by commercializing the roots and tubers production and processing businesses of smallholders, developing competitive market-driven and inclusive supply chains, and linking these to bigger markets. It pursued three important contribution claims to achieve this – with the district stakeholder forums (DSFs), farmer field forums (FFFs), good practice centres (GPCs) and Micro-Enterprise Fund (MEF).

- *Enhanced market linking.* DSFs would help develop the roots and tubers supply chains and link these to markets. Farmers and processors participating in the supply chains would gain better access to training and finance for investment and innovation, and be more able to commercialize their products and develop viable businesses.
- *Enhanced roots and tubers production.* FFFs would enable resource-poor farmers and seed producers to enhance their production and become commercial growers by adopting improved planting materials and technologies, and forming farmer organizations.
- *Enhanced roots and tubers processing.* Upgrading small roots and tubers enterprises into GPCs serving as demonstration and market hubs would attract small processors and expose them to high-quality processing operations using improved technologies and standardized equipment. This would help them enhance the quantity and quality of their production, obtain loans through the MEF to invest in these new technologies and standardized equipment, and develop profitable businesses.

Despite important differences in the context and quality of the two evaluations (discussed in the Reflections below), they both produced convincing evidence of the contribution of the projects to livelihood improvements as a result of increased access to markets created through the value chain development efforts. Although positive, the evidence also showed that these improvements were limited, fragile and susceptible to climate and market shocks, particularly for poor and vulnerable households and those in remote or marginalized areas. In both Ghana and Viet Nam, poor and vulnerable households ran considerable risks by engaging in value chains and accessing markets. These risks were left largely unmitigated due to: (a) inadequate links to markets and forecasting to avoid saturating or monopolizing local markets; and (b) inadequate poverty targeting to ensure support services and mechanisms were inclusive, sustainable and tailored to the needs of vulnerable groups and households (Van Hemelrijck, 2016c). To give an idea of the type of evidence that an evaluation using PIALA can produce, some of the key findings on the impact and contribution claims of the two projects are summarized below. Given the relative similarities of the two projects, some of the conclusions and recommendations go beyond the individual projects in question and have a wider relevance for IFAD investments elsewhere.

Key findings on the impact claims. Evidence from the DBRP evaluation in Viet Nam showed a positive change in household food security between 2007 and 2012. The change was quite small, given that most households were already food-secure prior to the project. But the evidence clearly showed a positive trend of poor households becoming middle-income, and middle-income households becoming well off. Among other poverty reduction programmes, DBRP contributed to increasing incomes. A key finding, however, was that while one third of poor households had experienced an increase in their incomes, roughly the same proportion had seen a decline. More than half of the non-poor households reported an increase in income, while fewer than one in five reported a decrease. Although the causes could not be inferred with great precision, the evidence pointed to critical resilience and sustainability issues. Risks and threats related to the market, jobs and climate had a negative impact on poor households' livelihoods, with environmental impacts being the major disruptive factor leading to job losses and making households more vulnerable. DBRP probably did not sufficiently target poor households and assist them in mitigating these risks and threats (IFAD and BMGF, 2014).

The RTIMP evaluation in Ghana also showed an improvement in household food security and significant increases in income across all socio-economic household categories. Livelihoods based on roots and tubers improved substantially between 2009 and 2013 in about 52 per cent of the supply chain areas, representing half of the country. Fifteen per cent of households increased their daily incomes to more than US\$2 (see Chart 1), which positively affected household access to food. Conversely, very weak or no improvements were found in supply chain areas where RTIMP mechanisms were dysfunctional or not in place. There were no households that made profits of more than US\$4 per day from roots and tubers, even if a relatively high percentage of households (61 per cent) had invested in roots and tubers businesses (see Chart 2). Producer prices for roots and tubers declined from 2013 onward due to weakening demand and insufficient market linkages, exacerbated by an economic downturn. Poor infrastructure and land tenure insecurity further limited the potential and incentives for smallholders to innovate and create added value (MOFA/GOG et al., 2015).

Chart 1: Change in distribution of the range of total household income in Ghana (2009-2014)

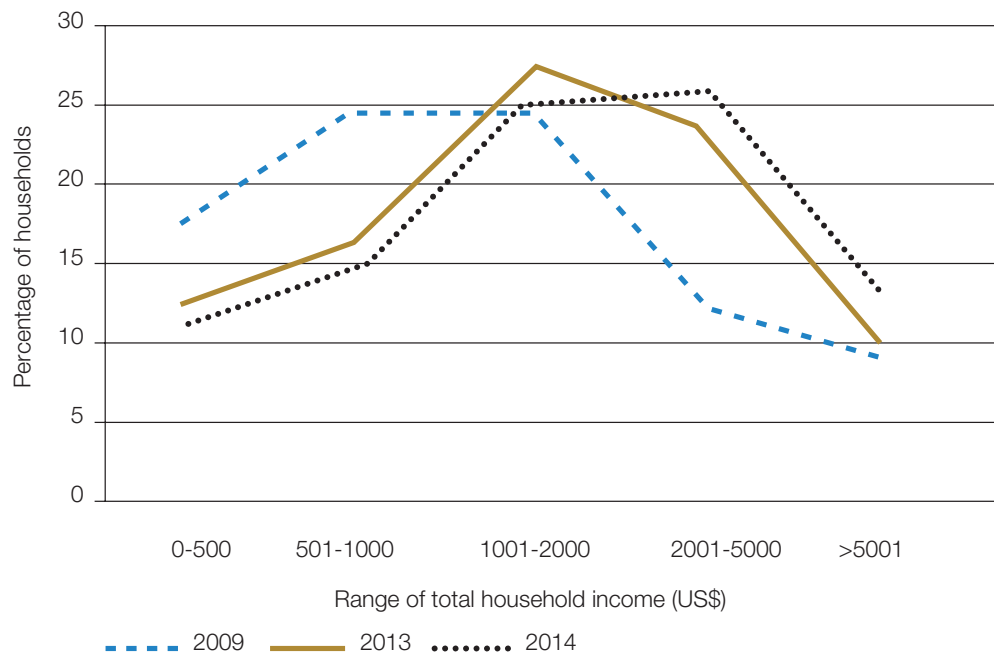
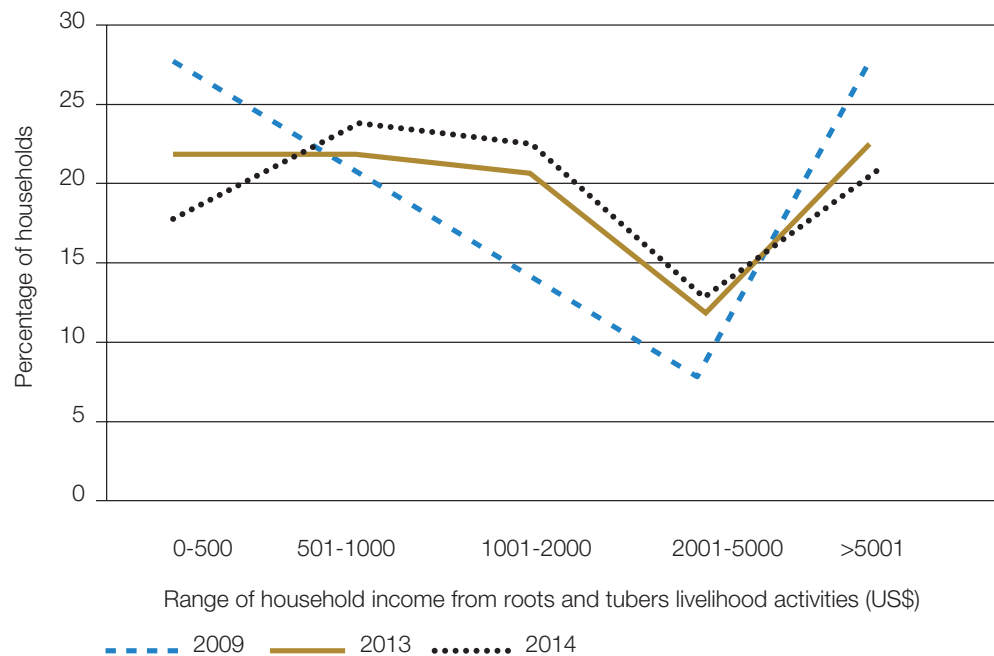


Chart 2: Change in distribution of the range of household income from roots and tubers in Ghana (2009-2014)



Key findings on the contribution claims. The DBRP evaluation showed that the proportions of women and men engaging in the different types of short value chains generally increased between 2007 and 2012. Livelihoods clearly diversified as a result of more opportunities. Women appeared more involved than men in developing new livelihood activities. Local trade increased and more jobs became available for landless households, but outmigration of young people also increased. The evidence suggested that these changes were brought about by more responsive institutions, improved local governance, increased access to diverse sources of credit with more favourable lending terms and interest rates, and other factors. The construction of roads, bridges and marketplaces, and the provision of training and services substantially contributed to the creation of market and job opportunities. Uptake of loans provided through the PPPs and CIG development fund remained limited, however. The relative benefits of training and services for poor and near-poor households also remained unclear. Most poor and near-poor households reported that they did not use the services, raising questions about the effectiveness of targeting and outreach (IFAD and BMGF, 2014).

The evaluation in Ghana showed that RTIMP had contributed to creating opportunities for farmers and processors by turning roots and tubers from subsistence crops into cash crops. The assumption that rural livelihoods and poverty status could be improved by commercializing roots and tubers smallholder businesses and developing competitive and inclusive supply chains proved to hold true, but only where the performance of district stakeholder forums (DSFs) and good practice centres (GPCs) was strong (12 per cent of the cases) in terms of supply chain linking, capacity-building, innovation and market creation, and reaching resource-poor farmers and processors. In other districts, investments in, and commercialization of, roots and tubers by smallholders remained limited and unsustainable because of the weak performance of the DSFs and GPCs, amplified by the economic downturn and other external factors. Farmer field forums (FFFs) clearly made a positive difference in 84 per cent of the supply chains by introducing new seed varieties and farming technologies that initially led to a boom in roots and tubers production across the country. But this success resulted in an influx into roots and tubers farming and triggered a spillover into processing (MOFA/GOG et al., 2015).

FFFs mostly targeted men and insufficiently reached women, despite the fact that they are generally more involved in cassava production than men. In contrast, the GPCs reached mostly women, as processing businesses were generally led by women. Processed volumes of cassava increased considerably in about 50 per cent of the supply chains, as a result of more women engaging in processing after receiving training at the GPCs. However, in 83 per cent of the cases studied, use of improved processing technologies and standardized equipment was limited, partly owing to the women having only limited investment capital. From 2010 onwards, only 15 per cent of the households received some sort of financial support to invest in their roots and tubers businesses. The Micro-Enterprise Fund (MEF) proved largely inaccessible because it required pre-investment without any short term capital return, posing a high risk particularly in the context of the economic downturn (ibid.).

Conclusions and recommendations. To reduce the risks caused by market saturation and monopolization, the RTIMP evaluation recommended a rethink of the DSF mechanism, making it into an independent forum for building inclusive supply chains and creating value, supported by a district administration with sufficient resources and capacity to

analyse markets, attract private investment, promote innovation and create opportunities for smallholder businesses. The DBRP tried to model this into a context-appropriate and formalized process of participatory socio-economic development planning (SEDP), with resources and capacities built and managed at the commune, district and provincial levels. However, the pro-poor short value chain approach was not always in evidence; there were few distinct synergies between the different value chain-related activities (credit, training and markets) and no clear links with enterprises. This shows that government-led multi-stakeholder forums for developing value chains and linking markets are in themselves insufficient to ensure inclusion and build resilience. The RTIMP evaluation provides evidence of the value of combining DSFs and GPCs for effective and inclusive market linking. In the few cases (12 per cent) where both performed well, livelihood improvements were found to be strong and resilient to shocks. In areas where one or both were weak or even dysfunctional, livelihood improvements were limited. In the context of DBRP, strategically linking the SEDP approach and GPC-like enterprises to specifically target resource-poor and vulnerable groups and households may offer scope for more synergistic effects (Van Hemelrijck, 2016c).

In both DBRP and RTIMP, there was an implicit assumption that all smallholders would want to commercialize their products, if given the opportunity to do so. It was also assumed that successful smallholders would be willing and able to help poorer and less successful households by offering them jobs and showing them how to operate in the market. Pairing them up and infusing some initial start-up capital into those likely to be more successful presumably would get this dynamic moving. But these assumptions were shown to be incorrect in both Ghana and Viet Nam. Not everybody can and wants to participate in the market, and not everybody can and wants to help those left behind. Adequate poverty targeting requires a good understanding of the profile, conditions and capabilities of the different groups and types of poor and vulnerable households so that targeting mechanisms can be tailored to reach them. Moreover, understanding how the local political economy excludes certain groups or households from accessing support and protection services is essential to adapting and adjusting targeting mechanisms in a timely manner. Proper feedback monitoring is also needed to keep track of the performance of these mechanisms (ibid.; MOFA/GOG et al., 2015).

In Viet Nam, many households did not participate in any major local organization or CIG. Understanding their reasons may help in identifying alternative strategies to strengthen and broaden the reach of such organizations (IFAD and BMGF, 2014). In Ghana, GPCs were supposed to become learning and market hubs for resource-poor farmers and processors. But unless there is a clear mutual business interest, farmers and processors with limited means do not go to these centres, and enterprises do not reach out to the farmers and processors. Initial funding is useful and often necessary to start the process, but this alone will not generate sustainable outcomes in the absence of a genuine business relationship. The power of a solid business relationship was shown in the few cases where GPCs functioned as open social-private profit centres. Where these GPCs proved to be profitable, this was largely due to their ability to innovate and create new market value, reach more farmers and processors (in particular women and young people) in the surrounding communities, and build trust between the various supply chain actors (MOFA/GOG et al., 2015). This successful GPC concept may be useful and effective elsewhere (also in Viet Nam) but to prevent the possibility of elite capture, the concept needs to be clearly defined, particularly its role in market linking and leadership. Performance and feedback monitoring mechanisms are also needed to help keep track of the quality and effectiveness of the business relationship with the surrounding communities, especially those that are poor in resources (ibid.; Van Hemelrijck, 2016c).

Reflections on the approach in evaluations that piloted PIALA⁸

As an action research project, the successes and problems of the pilots were debated and documented in great detail with researchers and project stakeholders within IFAD and externally. Insights from the Viet Nam pilot greatly strengthened the work in Ghana. This section summarizes the most relevant learning points from the three phases of the evaluation: creating ownership and deciding on the evaluation scale and scope in phase 1; deciding on the counterfactual, maintaining independence, contextualizing poverty analysis, and dealing with power and bias in phase 2; and deciding on the scale and level of engagement in sense-making, and linking impact to organizational performance in phase 3 (see Figure 1 for the three phases).

First phase – focusing and framing the evaluation

Creating ownership of the evaluation. Ownership implies that the evaluation is wanted, legitimized and supported by a shared sense of responsibility among stakeholders. This facilitates greater uptake of and learning from evaluation findings and recommendations (Burns and Worsley, 2015; Patton, 2011). Ownership is created by sufficiently engaging stakeholders in the process of reconstructing and visualizing the programme's ToC to frame and focus the evaluation.

In Viet Nam, insufficient time and budget was spent on this process, which affected the rigour and inclusiveness of the approach throughout the evaluation. Project management and steering committee members were consulted and to some extent engaged in discussions on programme logic and expectations for the evaluation. They were also involved in designing the evaluation (particularly the M&E team) to ensure that it would build on what was already known. But no space was created for them to engage in the reconstruction of the ToC as the basis for collectively determining the focus of the evaluation and agreeing on the way data would be linked to assess the programme's contribution claims. In fact, the evaluation design took place separately, before the ToC process, and it was guided by generic questions taken from the PIALA initiative strategy, rather than questions designed to investigate specific assumptions underlying the DBRP's ToC. As a result, the evaluation scope ended up being too wide and its focus unclear, making it difficult for the researchers to relate the evidence back to the ToC and to be sufficiently precise in their causal analysis. Furthermore, limited ownership also hindered stakeholders' ability to critically engage in sense-making and analyse contributions in the last phase of the evaluation (IFAD and BMGF, 2013a).

8. This section draws on the reflection reports written by one of the authors, Adinda Van Hemelrijck, as part of her doctoral study (see Van Hemelrijck, 2016b) and the PIALA initiative (see IFAD and BMGF, 2013a, 2015).

Armed with these insights, the ToC process was made a priority and a key deliverable in Ghana. Researchers were required to lead this process and organize a design workshop with programme management, M&E staff and other national stakeholders to consolidate all inputs into the process and determine the frame and focus of the evaluation. This investment in a robust and collaborative ToC process bore fruit and laid the foundation for attaining greater quality throughout the entire evaluation, resulting in stronger evidence and stronger ownership of the evaluation processes and findings (IFAD and BMGF, 2015).

Deciding on the scale and scope of evaluation. PIALA's mixed-methods approach pursues depth and breadth; it seeks depth through a focused participatory inquiry on "open systems" (such as value chains) and breadth through a representative household survey and participatory inquiry within these systems. The sample size of the open system determines the scale of the evaluation; households for survey and intended beneficiaries for participatory inquiry are subsampled from within these open systems. The larger the scale, the more relevant the findings for national and global policy engagement will be, but the more challenging it will be to produce rigorous evidence. Mixed methods, particularly participatory methods, become more onerous when used on a large scale and when resources are tight. Nevertheless, depth of analysis does not have to be sacrificed for breadth of coverage (or vice versa), if sufficient capacity and motivation to deliver quality is present. In contexts where local research capacity is weak, ambitions in terms of scale and scope need to be lowered and sufficient resources made available for training, coaching and supervision (Van Hemelrijck and Guijt, 2016). For example, the later PIALA evaluation of local inclusive governance for building resilience in Myanmar was conducted with inexperienced researchers from the villages. The design was kept simple, and the scope and scale of the evaluation limited. Sufficient data was obtained to reconstruct the causal chain, cross-check findings, and compare and aggregate them for the entire project. This resulted in robust evidence showing the value and sustainability of the governance model and its contributions to resilience. Important issues were raised about point-of-exit, government responsiveness, collective power, leadership and gender (Van Hemelrijck, 2016a, 2017).

Whereas scale is determined by the size of the sample of systems, scope refers to the span and complexity of the system and the various project components and mechanisms to be covered. Three options are possible for designing the evaluation: full scope–limited scale, limited scope–full scale or full scope–full scale. When choosing a full scope–limited scale design, the emphasis is on learning about the project's total contribution to impact in selected cases under specific conditions. Here the ToC approach is most useful for systemic learning with key stakeholders. Fieldwork and analysis are less resource-intensive as the samples are relatively small. But unless the project itself is implemented on a limited scale, it will not be possible to generalize evaluation findings for the entire population. The evaluation is therefore less useful for influencing policy decisions. When choosing a limited scope–full scale design, the purpose is to learn about the effects of one or two particular aspects or mechanisms of the project. The ToC is not strictly necessary for a narrow study. Skipping this step will save time and money, but there is a risk of omissions in the systemic understanding, leading to flawed conclusions. Components are studied in isolation, which does not allow analysis of systemic interactions (IFAD and BMGF, 2015). For example, a cost-effectiveness study of farmer field schools in Ghana had recommended scaling these up because the high adoption of new technologies was considered sufficient evidence of the success of this

mechanism. The PIALA evaluation showed, however, that in a period of weak economic growth, this success contributed to market saturation, which negatively affected livelihoods across the country.

In Viet Nam, a full scope–limited scale design became the *de facto* choice, but with a disparate scale for different methods. Only a subsample of villages was included in the participatory research, drawn from a wider sample of villages where household surveys were conducted. The asymmetry in the two data sets caused problems for subsequent data linking: while participatory data on causes and contributions were case-specific and limited to a few villages, survey data on impact applied to a much wider sample and did not relate to specific cases or villages (IFAD and BMGF, 2013a).

In Ghana, a conscious choice was therefore made to use all methods in the same sample on the same scale. Before any procurement or design work was started, the three design options were discussed with clients and commissioners (including the Ministry of Food and Agriculture and the IFAD country office), giving them a good understanding of the value for money to be expected from each option. As the follow-up Ghana Agricultural Sector Investment Programme (GASIP) was expected to scale up some of the RTIMP mechanisms, the evaluation was important for both reporting and learning. The commissioners therefore chose the most substantive and expensive design option: full scope–full scale. This required six weeks of fieldwork, compared with only two in Viet Nam. The budget was relatively tight compared with Viet Nam, given the larger scale and scope, but quality was maintained by working with a highly competent and motivated research team (IFAD and BMGF, 2015; MOFA/GOG et al., 2015).

Second phase – collecting and linking data

Deciding on the counterfactual. Mainstream impact evaluation assumes that comparative analysis of evidence from treated and non-treated locations is both feasible and necessary if conclusions that can be generalized are to be reached about impact on rural household poverty. In most “real world” evaluation contexts, however, it is very difficult and costly, even impossible, to accurately assign locations to specific interventions and identify control groups (Bamberger et al., 2012). A particular challenge arises in cases of unexpected or uncontrolled project expansion and/or spillover, combined with a high causal density of other interventions and influences, which makes it become difficult to separate project from non-project localities and find the right matches (Woolcock, 2009). In addition, the principal unit of analysis from which to subsample households is not a clear physical unit with well-defined boundaries, such as a community or village, but an open system. Identification of matching control units thus requires proper definition of this “system”, and additional fieldwork needs to be done (before the evaluation) to identify matching systems and households within these systems, all of which adds to the cost.

Such an investment may be worthwhile when the focus of evaluation is on learning about the difference a specific intervention or mechanism made on households, without needing to dig deeper to find out why it worked in some contexts and not in others, and whether the observed effects are likely to be sustainable or not. The value of PIALA is that it can assess, explain and debate project contributions to systemic change which have had an impact on households. The primary focus is on understanding systemic change, not on measuring the net impact an isolated intervention has had on households.

In the Viet Nam pilot, we assumed that comparative analysis of treated and non-treated units would be possible and necessary to draw conclusions regarding household-level impacts, and that the village constituted the best proxy unit for investigating the short value chains developed by the programme. These assumptions were flawed and compromised on analytical rigour, making it difficult to generalize findings – for three important reasons. Firstly, without a clear definition and identification of value chain systems, and without sampled proper proxies based on such a definition, it was difficult to link the data on the various processes and changes observed in capacities, institutions and livelihoods, and relate them to specific value chains. Secondly, the matching of treated and non-treated locations was based on variables applicable to the village and not to the value chains, making it difficult to relate the observed changes and outcomes and impacts back to the project. Thirdly, high heterogeneity in project delivery and incoherence in its value chain linking efforts, further conflated by the high causal density of other programmes and influences in the villages, made it even more difficult (if not impossible) to identify credible factual and control units without a costly pre-evaluation study (IFAD and BMGF, 2013a).

Learning from this, more work was invested in understanding and defining the system sampled in Ghana. At the evaluation design workshop, it was decided not to channel resources into identifying and investigating control groups, but instead to concentrate on analysing local supply chain systems of commodities promoted by RTIMP. These supply chains were made up of loose catchment areas comprising clusters of communities of smallholders supplying raw products to a small-scale enterprise or industrial off-taker, known as the supply chain leader, who manufactures higher-value products for bigger markets. The chains were not homogeneous: they interacted and overlapped, and were often different from what they appeared to be on paper. In some cases, suppliers produced different crops supplying different buyers and markets; in other cases, multiple chains intertwined. Ensuring the evidence collected on these systems remained comparable required creativity and coordination (IFAD and BMGF, 2015).

In addition, no reliable supply chain-specific lists of households and intended beneficiaries were available. Proper identification and matching of control units and sampling of households within these units would have required extensive pre-evaluation fieldwork. Comparison with classic control groups, therefore, was replaced with a configurational analysis method using systemic heterogeneity as the basis for the counterfactual analysis. Supply chains with different systemic configurations of treatment and conditions were randomly sampled (with probability proportional to size) from the supply chain populations for the four commodities, ensuring the samples also included supply chains with dysfunctional or absent RTIMP mechanisms to serve as a counterfactual (MOFA/GOG et al., 2015; Van Hemelrijck, 2016b, 2017).

Maintaining independence. Balancing the cost-efficiency of fieldwork with independence is a common challenge in impact evaluation. Field mobilization of participants in the evaluation is best undertaken independently from project management to avoid bias.⁹ When participants suspect the evaluation is not independent, they are more likely to over- or under-report. On the other hand, they are likely to distrust outsiders who are not authorized and formally introduced by their leaders. Thus, for the researchers to organize fieldwork at scale and mobilize participants without any help from the project, they need to be skilled in

9. According to the OECD (2010, p.14), “independence” implies an evaluation process that is transparent, independent from project management and free from political influence or organizational pressure.

logistics, have knowledge of the local areas and customs, and be able to obtain authorizations and introductions in ways that do not affect their independent status. In contexts where this is not possible, strong facilitation skills are needed to minimize undue influence or interference (Van Hemelrijck, 2016c).

The challenges encountered in Ghana were quite different from those in Viet Nam. Yet in both cases, it was observed that participants' confidence in the independence of the researchers gave them the freedom and the space to form and express their own views and critically engage in collective discussions. In Viet Nam, fieldwork cannot be conducted without government permission and control. In evaluations of government projects, logistics are taken care of by the government. Hence in the DBRP evaluation, local transportation and mobilization was organized by local officials and project staff. This made the work highly efficient but challenged the researchers' independence. Although collaborative and supportive, local leaders and project staff were omnipresent during fieldwork. The researchers managed to keep them at a distance and safeguard the privacy of focus groups. Local project staff and officials were more genuinely engaged as the evaluation unfolded, however, and had a greater sense of ownership of the results (IFAD and BMGF, 2013a).

In Ghana, the researchers took care of the transportation and mobilization themselves and without prior notification or engagement of local officials, which allowed for much greater independence. Staff and officials were present at discussions only when invited. The downside was that they were more resistant to the evaluation and less engaged in the discussions. The scale of the fieldwork, the remoteness of the communities and the long distances to be travelled over often very poor roads made the exercise quite onerous. It was also difficult to find safe and trusted locations for convening people from different communities. Independence comes at a considerable cost in countries such as Ghana, yet it does contribute to the rigour and inclusiveness of the evaluation (MOFA/GOG et al., 2015).

Contextualizing poverty analysis. To say something about how a project has influenced poverty, data on such influences and on poverty must be linked. Poverty needs to be defined in ways that are relevant to the context and conditions of villagers in the project area. In both the Viet Nam and the Ghana pilots, a method known as participatory ranking was used to identify locally relevant indicators of wealth and wellbeing and analyse changes in relative poverty status. In both cases, the method helped to create a shared understanding among the participants of relative wealth and wellbeing, before turning to a causal flow mapping of changes in their wealth and wellbeing. It was also possible to cross-check and link the participatory data with the household survey data (IFAD and BMGF, 2015).

The wealth and wellbeing characteristics obtained from these exercises were not used as indicators when assessing changes in poverty status through the household survey because this would have required participatory data collection prior to the evaluation as an input for designing the household survey – again adding to the cost. It is unclear whether, and to what extent, this could have generated more rigorous findings on poverty impact, thus justifying the extra investment. In Viet Nam, the survey focused only on IFAD's generic poverty indicators and used the income-based absolute poverty categories of the Vietnamese government. These proved inadequate for assessing changes in poverty status in the project area (ibid.).

As a result, in Ghana greater effort was made to ensure the household survey questionnaire was attuned to reality on the ground. Poverty characteristics were selected from the Ghana Living Standards Survey 2009-2014, which corresponded with IFAD's poverty indicators. No major differences were found between the characteristics obtained from the participatory ranking exercise and those used by the household survey. Poverty status was determined by first applying a proxy means test to assess each household against each of the selected characteristics. The data collected through the household survey were then used to compute the categories of poverty status (applying a principal components analysis). Greater rigour could have been achieved in the findings on poverty distribution and impact if the questionnaire had gone into more detail about household characteristics. But more lengthy surveys would have added to the cost, while also increasing the risk of fatigue and gaming among respondents and researchers (Chambers, 2008). Therefore, in both Vietnam and Ghana, the surveys were kept to 20 minutes and participatory group discussions were kept within a two-hour limit, precisely to avoid such research fatigue (IFAD and BMGF, 2015).

Thus, instead of spending more resources on collecting and analysing participatory poverty characteristics prior to the evaluation, or on collecting more fine-grained data on household characteristics to identify poverty categories, in Ghana the choice was made to keep the poverty analysis short in order to create more room for participation in ways more meaningful to the participants. In particular, participants found the group-based causal flow mapping exercises useful, as the exercises helped them recall and understand the changes from a systemic perspective. It also equipped them to engage in collective sense-making of evidence and discuss contributions and responsibilities with other stakeholders (see the last learning point – the third phase – in this section). It was assumed that these interactions and discussions enable people to “see the system more effectively from different perspectives” (Burns and Worsley, 2015); contribute to building competencies to better navigate the system and project agency; and ultimately empower them (Gaventa, 2004; Merrifield, 2002; Van Hemelrijck, 2013).

Dealing with power and bias. All methods are susceptible to bias. Participatory methods, however, are considered more vulnerable than traditional survey-based methods as they collect perceptions, meanings and interpretations instead of numbers. Yet numbers are also generated by surveys designed and conducted by people with value judgements (Camfield et al., 2014; Copestake, 2013; White and Phillips, 2012). Survey questionnaires tend to reflect the designers' assumptions, while qualitative interviews and participatory inquiries give more room to the assumptions of the people being researched. Both types of methods may result in desirability or courtesy bias if those being researched tell the researchers what they think is wanted. As Sarah White (2015: 138) points out in her review of a large-n RCT evaluation of girls' empowerment in Bangladesh: “hypothetical questions are susceptible to ‘desirability bias’ in which respondents give the answer they believe the researchers wish to hear.”

To overcome potential bias, the PIALA evaluations used mixed methods with different groups, enabling extensive cross-checking of different methods and perspectives at scale. Yet scale may create new biases, as it tends to instrumentalize and thus de-politicize the research context and relationships, thereby increasing the risk of power-blindness and bias (Hickey and Mohan, 2004; Mosse, 2001). In the two pilots, attempts were made to avoid this by arranging and facilitating the group processes to: (a) reduce power imbalances, for example by making sure that power holders remain in the minority in mixed groups; and

(b) enable people to see how data is constructed and indicate where things are flawed by using visual tools, such as causal flow mapping and matrix scoring. There is, of course, always a danger of more powerful participants dominating the discussions, although good facilitation can overturn this. Rigour derives from sharp observation of motives and interactions and continuous cross-checking of different sources (Chambers, 2015).

In Viet Nam, this was quite challenging for the researchers, whose background and expertise was principally quantitative. They struggled with triangulation as a way of compiling a multi-perspective picture that forestalls a single dominant truth. They also had limited awareness of their own position and influence, and found it difficult to uphold a daily practice of critical self-reflection on quality and process. In Ghana, by contrast, the researchers had mixed backgrounds and substantial experience in participatory research, while the research assistants were from the areas where the fieldwork was conducted and thus were familiar with the local languages, cultures and political economies (IFAD and BMGF, 2015). This helped to avoid bias caused by interpretation (Camfield et al., 2014).

Third phase – analysing contributions

Deciding on the scale and level of engagement in sense-making. Participation in evaluations is purely extractive if findings are not shared with those who have given their time and experiences and no opportunity is created for them to contest and debate the findings (Gaventa, 2004; Hickey and Mohan, 2004). Using PIALA's sense-making model, six village-level workshops with 180 participants and one provincial workshop with 100 participants were organized in Viet Nam. In Ghana, 23 district workshops with 650 participants and one national workshop with over 100 participants were organized. Workshop participants were selected from among the research participants (IFAD and BMGF, 2015).

There were two main outcomes. First, an additional layer of detail and confirmation of evidence was obtained from the cross-validations held in the local and aggregated workshops, adding to the rigour and validity of the findings.¹⁰ Second, a strong sense of shared ownership of both the evidence and the findings was created, contributing to the evaluation's inclusiveness and its empowerment value. Having participated only in data collection, people arrived at the workshops knowing and owning little. In Ghana in particular, they left the workshops with a more comprehensive picture of the systemic changes and issues that the evidence had revealed and stakeholders' different perspectives on these. Indeed, the end-of-workshop survey and reflections revealed a high level of satisfaction among participants and knowledge and insights gained, and that they would be useful for future individual or collective action. Critical to this success was the scale of the workshops and the way in which they were designed and facilitated. Successful sense-making at scale requires capacity and resources. When operating on a shoestring budget or in a context of low capacity, the number and size of workshops may need to be reduced at the expense of rigour and inclusiveness¹¹ (Van Hemelrijck and Guijt, 2016).

10. "Validity" is understood as the extent to which findings are based on robust evidence and correspond to the reality of all the populations of the project being evaluated.

11. This was the case in the evaluation of local inclusive governance in Myanmar. Because of the limited capacity of the village researchers and the limited resources, sense-making had to be downscaled and local workshops skipped. Compared with Ghana, where over 70 per cent of all research participants also participated in local and national sense-making workshops, only 5 per cent (or 50 participants) did so in Myanmar. In short, many local people were left out of the debates (Van Hemelrijck, 2017).

A truly participatory sense-making process implies equal and active engagement of all stakeholders. Dynamic environments were created, long presentations discouraged and different sorts of evidence made available in accessible (including visual) formats. Group work was designed to make people feel “listened to” rather than just “talked at” (Newman, 2015). Intended beneficiaries constituted over 30 per cent of the participants at the provincial and national workshops and 70 per cent at the local workshops, which gave them sufficient weight in debates with decision makers and service providers. At the local workshops, the proportion of intended beneficiaries needs to be larger because they are the project’s primary target group at the local level, while at the aggregated level the primary targets are policymakers and service providers. Lessons learned from the Viet Nam workshops helped improve the model for Ghana.

In Viet Nam, discussions took place in mixed stakeholder groups and at plenary sessions, which did not give the beneficiaries enough space to collect their thoughts and speak with confidence. Learning from this, participants in Ghana first worked with groups of peers (organized around the part of the ToC representing their link in the supply chain, e.g. production). In Viet Nam, the reconstruction of the causal flow was done at plenary sessions, which again did not leave sufficient room for beneficiaries to engage. In Ghana, this was done in small mixed groups (organized around geographic areas, in which beneficiaries presented the findings from their supply chain groups. The plenary discussion took place only on the second day in a fish bowl set-up, in which most of the discussants were beneficiaries. This provoked an animated discussion with the bankers around the inaccessibility of the Micro-Enterprise Fund (see key findings on contribution claims in the section on what emerged from the evaluations that piloted PIALA) (Van Hemelrijck, 2016c).

Linking impact to organizational performance. A major challenge of impact evaluation is to make learning and feedback actionable for management and staff. Impact evaluation, it is often argued, tends to give somewhat “clueless feedback” about a project’s contribution to impact (Earl and Carden, 2002, p. 519): it does not indicate how things could or should have been managed and implemented differently in order to attain greater impact. While process and performance evaluation falls short of assessing impact, impact evaluation fails to draw lessons about organization and operations. This was considered an important limitation in the RTIMP evaluation in Ghana, where the IFAD country team had expected PIALA to make a difference. In the words of the IFAD country programme manager: *“We may have designed the right thing (linkages or matching grant funds) but due to implementation constraints, performance was weak and overall results were weak as well. Many of the shortcomings were documented in the supervision reports. Looking ahead, I suggest developing the PIALA tools further to allow disaggregation of the conclusions on the approach, particularly conclusions about implementation performance.”*¹² Many of the shortcomings were indeed documented in the supervision reports that were part of the evaluation’s desk review. But the evaluation deliberately shied away from detailed analysis and judgment of operational performance in order to avoid shifting attention from the change processes affecting impact to the organizational processes affecting performance.

There were two main reasons for this. First, conducting a rigorous combined performance evaluation and impact evaluation on a large scale would have required a much bigger budget. Because they are resource-intensive, end-of-term evaluations often take a case-based approach to assessing outcomes and impacts (using a small-scale or a small-n design), in order to leave enough room for assessing organizational performance. In-depth inquiries in such evaluations involve interviews with staff and partners and a review of M&E data and

12. Personal communication from Ulaş Demirag, at the time IFAD Country Programme Manager in Ghana.

supervision reports directly related to the field data collected on how well project mechanisms functioned in a small number of selected cases. The challenge is to rigorously link data when reconstructing the causal chain as the basis of causal analysis (from impact to organizational performance). Within-case analysis methods, such as process tracing, are used to arrive at a detailed descriptive reconstruction of how events unfolded over time as the basis for causal inference (Befani et al., 2015; Collier, 2011). However, such a detailed case-based approach is unable to reach generalizable conclusions on outcomes and impacts for larger populations (which is the type of evidence generally needed for influencing policy).

Second, project managers may feel threatened by an evaluation that focuses on organizational performance, particularly in complex environments where there is less control over results and similar implementation processes may lead to different outcomes. This fear may hinder solid debate and collaborative learning about impact together with other stakeholders. By focusing on impact and seeking to understand contributions to impact alongside many other influences, the fear of failure can be partially sidestepped (Eyben et al., 2015); though the consequence is that responsibility for failure can, equally, be sidestepped (Gorgens and Kusek, 2009; Kusek and Rist, 2004). Addressing this limitation would require not only another layer of data collection and analysis but also another separate layer of sense-making further down the causal chain that is more narrowly focused on organizational performance and inciting action learning with the implementing partners.

The challenge would be to include this extra layer of data collection, analysis and sense-making in a way that makes it possible to link impact assessment to organizational performance on a large scale. This would allow findings to be generalized for large populations. PIALA could potentially do this in a cost-effective manner if it were used as an overarching longitudinal approach supported by an appropriate learning architecture and embedded in project design.

Lessons learned and critical next steps for PIALA

The PIALA initiative sought to design and test an approach that would complement those already being used and be useful to IFAD and its partners for learning, reporting and advocacy. The two pilots have shown that a participatory and systemic impact evaluation approach such as PIALA can produce rigorous quantitative and qualitative evidence of impact on a scale useful for policy debate and for learning with partners and stakeholders in contexts where traditional counterfactuals do not work. Although more piloting in different contexts is needed to address all the remaining questions and weaknesses, the pilots did show the potential of the approach for supporting IFAD's self-evaluation system and knowledge agenda.

The results of the two pilots confirm that PIALA can indeed produce rigorous evidence. The pilots also brought important common issues to the surface, such as weak and gender-insensitive poverty targeting and inadequate market forecasting, which resulted in unsustainable livelihood improvements susceptible to climate and market shocks (see the section on what emerged from the evaluations that piloted PIALA). This demonstrated PIALA's potential for offering evidence for global policy and building knowledge to improve IFAD engagement.¹³ It may be possible to aggregate evidence across interventions with comparable ToCs and intervention logics for conducting systematic reviews. The pilots also demonstrated PIALA's potential value in empowerment, or how engaging rural populations in assessing and debating the evidence of impact can potentially contribute to enhancing impact even ex post.¹⁴ This not only supports IFAD's underpinning principle of promoting empowerment (IFAD, 2016, p.18); it is also particularly relevant in the context of the 2030 Agenda for Sustainable Development, which calls for fundamental systemic changes, inclusiveness ("leaving no one behind") and sustainability, in addition to effectiveness (Van Hemelrijck, 2016c).

Compared with how much theory-based mixed-methods impact evaluations normally cost in countries such as Viet Nam and Ghana, the PIALA studies were done on a shoestring: US\$90,000 in Viet Nam and US\$234,000 in Ghana. (These amounts do not include the investment costs of PIALA design, innovation and meta-inquiry – US\$100,000 in Viet Nam and US\$60,000 in Ghana.) The cost of the PIALA evaluation of RTIMP includes an assessment of three programme components and multiple mechanisms in 30 districts across the entire country.¹⁵ It also includes a statistical survey of 900 households and a participatory inquiry with over 1,300 participants, nearly half of whom were women. The exercise was completed

13. These commonalities probably would have been much clearer and lessons more comparable if the Viet Nam pilot had had the more mature design of the Ghana pilot. But since Viet Nam was the first pilot, the potential for learning is evident.

14. Statement from Ulaş Demirag, at the time IFAD Country Programme Manager in Ghana.

15. This includes only a small part of the cost of international consultants – that directly related to evaluation design and supervision (i.e. not innovation and meta-analysis).

Table 2: PIALA budget and time requirements

	Ghana		Myanmar	
	Budget (US\$)	Time (weeks)	Budget (US\$)	Time (weeks)
Phase 1 Desk review, ToC process, design workshop	18 000	3	8 000	1
Phase 2 Sampling, training, field-testing, data collection, local sense-making	152 000	8	23 000 (No local sense-making)	7
Phase 3 National sense-making, analysis, reporting	64 000	8	31 000	4
Total	234 000	19	62 000	12

in less than five months. By contrast, the estimated budget for a one-year RCT study in an IFAD-funded project similar to RTIMP in northern Ghana was also around US\$200,000 – but it covered only one subcomponent and was conducted in only eight districts.

Also worth citing is a subsequent Oxfam GB Effectiveness Review (Van Hemelrijck, 2016a) that adopted PIALA to evaluate the impact and sustainability of local inclusive governance on resilience in the Dry Zone in Myanmar 18 months after project exit. This evaluation cost US\$62,000 and took only three months to conduct. It had a scope of only one causal claim and one mechanism, and was conducted in two townships that were a four-hour drive apart. It involved a participatory inquiry involving 1,030 participants (nearly half of whom were women) across a sample of 21 villages, but it did not include either a statistical household survey or a local sense-making exercise. The sample reflected the different levels and configurations of membership organizations functioning 18 months after project exit, determined through a rapid survey of all 64 project villages. The evaluation was conducted with people from local villages who had no previous research experience, and with very limited support from the local partners, given that the project was already closed (Van Hemelrijck, 2017). The IFAD pilots and the Oxfam GB Effectiveness Review demonstrate PIALA's ability to adapt to different scales, budgets and capacities (see Table 2).

Trade-offs occur in every evaluation that aims to create greater value with limited resources. But the IFAD pilots have shown that these trade-offs can be turned into win-wins by carefully considering how rigour and inclusiveness can reinforce each other and what would be the potential loss in value for money if one were to be prioritized over the other (Van Hemelrijck and Guijt, 2016). Inclusiveness and rigour reinforce each other when participatory processes and methods are employed on a large enough scale and thoughtfully designed and facilitated in ways that forestall power and bias (or dominance of a single truth), while enabling participants to take ownership of the findings (Burns and Worsley, 2015; Van Hemelrijck, 2016c).

Critical to quality delivery at scale are the need for investment both in researchers' capacity and the further development of the PIALA approach. Investing in building such capacity helps to reduce the cost over time, while enhancing the value and uptake of impact evaluation. One way to build capacity will be to invest in further action research and testing of the approach through new PIALA evaluations. Much remains to be learned about how to do it well. Already, the results of IFAD's initiative have attracted attention among evaluators and commissioners and generated a demand for practical guidance and training in the international development community. An online "living" guide could help evaluation commissioners and practitioners find out whether, when and how PIALA could be useful and adaptable to their specific evaluation context and needs. Such a guide would be able to expand as new evaluations adopting PIALA become available and new insights emerge as to how their design and management can be adapted to different contexts, conditions and needs. A standardized training package could help build the capacity of commissioners, managers and practitioners.

For IFAD, its partner governments and other development agencies, optimizing value for money could imply using PIALA as a longitudinal approach built into project and programme design. This could provide an opportunity to link impact evaluation to organizational performance at a large scale and to invest in building local research partnerships and capacity for impact evaluation. This, in turn, could bring greater continuity and consistency to the impact learning agenda of IFAD-supported projects. It would also make it feasible to embed other methodologies – such as an RCT, a governance assessment method or a targeting and performance feedback tool. PIALA's ToC approach is particularly helpful for deciding, together with project partners, which methods to use for investigating specific project mechanisms or causal links at any given moment in time.

Depending on the requirements of the different methodologies, a mixed-counterfactual approach could be designed using PIALA's configurational method for the analysis of the "systems" populations and a classic counterfactual method for comparing factual and control populations within the "systems" populations (which would require identification of comparable non-treated entities and proper construct and definition in project design). Participatory sense-making workshops could be organized at the baseline, mid-term and end-term of the project and would form the basis for establishing a multilayer and multi-stakeholder learning architecture around impact and contributions, linked to project management and performance. Such a learning architecture would enable collaborative action and learning with project stakeholders on a larger scale, leading in turn to greater ownership of the change processes, and thus presumably wider adoption of the knowledge being produced. At the same time, it would facilitate the translation of findings and reflections into future project design, and provide evidence to inform policy decisions. Ultimately, it could make a significant contribution to improved development results, and to greater and more transformative impact.

References

- Abebe, D., Catley, A., Admassu, B., and Bekele, G. (2009). Using Participatory Impact Assessment (PIA) to inform policy: Lessons from Ethiopia. In: *Farmer First Revisited. Innovation for Agricultural Research and Development*. IDS.
- Aus, J. P. (2005). *Conjunctural causation in comparative case-oriented research*, (28).
- Bamberger, M. (2012). *Introduction to mixed methods in impact evaluation* (Guidance note No. 3). InterAction, Rockefeller Foundation.
- Bamberger, M., Rugh, J., and Mabri, L. (2012). *RealWorld Evaluation*. SAGE.
- Bamberger, M. and White, H. (2007). Using Strong Evaluation Designs in Developing Countries: Experience and Challenges. *Journal of MultiDisciplinary Evaluation*, v4 n8.
- Beach, D., and Pedersen, R. B. (2013). *Process-tracing methods: foundations and guidelines*. University of Michigan Press.
- Befani, B., D'Errico, S., and Giuliani, A. (2016). *Clearing the fog: new tools for improving the credibility of impact claims*. IIED Briefing Paper.
- Befani, B., Barnett, C., and Stern, E. (2014). Introduction – Rethinking impact evaluation for development. *IDS Bulletin*, 45(6).
- Befani, B., and Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *IDS Bulletin*, 45(6).
- Befani, B., Ramalingam, B., and Stern, E. (2015). Introduction – Towards systemic approaches to evaluation and impact. *IDS Bulletin*, 46(1).
- Burns, D. (2014). Assessing impact in dynamic and complex environments: Systemic action research and participatory systemic inquiry. *CDI Practice Paper*, (8).
- Burns, D., and Worsley, S. (2015). *Navigating complexity in international development: Facilitating sustainable change at scale*. Practical Action Publishing.
- Burns, D., Oosterhoff, P., Bharadwaj, S., Mohan Raj, A., Nanda, R., and Narayanan, P. (2016). Using participatory statistics to examine the impact of interventions to eradicate slavery: Lessons from the field. *CDI Practice Paper*, (16).
- Camfield, L., Duvendack, M., and Palmer-Jones, R. (2014). Things you wanted to know about bias in evaluations but never dared to think. *IDS Bulletin*, 45(6).
- Catley, A., Burns, J., Abebe, D., and Suji, O. (2008). *Participatory impact assessment. A guide for practitioners*. Feinstein International Center, Tufts University.

- Chambers, R. (2008). *Revolutions in development inquiry*. Earthscan.
- Chambers, R. (2015). Inclusive rigour for complexity. *Journal of Development Effectiveness*, 7(3).
- Collier, D. (2011). Understanding process tracing. *PS: Political Science and Politics*, 44(4), 823–830.
- Copstake, J. (2013). *Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches* (Draft 18 Oct 2013). Centre for Development Studies, University of Bath.
- Earl, S., and Carden, F. (2002). Learning from complexity: the International Development Research Centre's experience with outcome mapping. *Development in Practice*, 12(3 and 4).
- Eyben, R., Guijt, I., Roche, C., and Shutt, C. (2015). *The politics of evidence and results in international development: playing the game to change the rules?* Rugby, Warwickshire: Practical Action Publishing.
- Funnell, S. and Rogers, P. (2011). *Purposeful program theory: effective use*. John Wiley & Sons.
- Gaventa, J. (2004). Towards participatory governance: Assessing the transformative possibilities. In *Participation: from Tyranny to Transformation? Exploring New Approaches to Participation in Development*. Zed Books.
- Gorgens, M., and Kusek, J. Z. (2009). *Making monitoring and evaluation systems work: A capacity development toolkit*. World Bank.
- Guijt, I. Woodhill, A.J. (2002). *Manging for impact in rural development: a practical guide for M&E*. IFAD.
- Hickey, S. and Mohan, G. (2004). *Participation – From tyranny to transformation: exploring new approaches to participation in development*. Zed Books.
- Holland, J. (2013). *Who counts? The power of participatory statistics*. Practical Action Publishing.
- Holland, J., and Garbarino, S. (2009). *Quantitative and qualitative methods in impact evaluation and measuring results*. Issue Paper.
- IFAD. (2005). *Results and impact management system. Practical guidance for impact surveys*.
- IFAD. (2011). *IFAD Strategic Framework 2011-2015*. Enabling poor rural people to improve their food security and nutrition, raise their incomes and strengthen their resilience.
- IFAD. (2012). *Methodologies for impact assessments for IFAD IX* (Note to IFAD's Executive Board Representatives).
- IFAD. (2016). *IFAD Strategic Framework 2016-2025*. Enabling inclusive and sustainable rural transformation.
- IFAD and BMGF. (2013a). *Improved learning initiative for the design of a participatory impact assessment and learning approach (PIALA): Insights and lessons learned from the reflections on the PIALA piloting in Vietnam*. IFAD.
- IFAD and BMGF. (2013b). *Participatory Impact Assessment and Learning Approach: A Research Strategy for IFAD's Improved Learning Initiative* (Internal Document). IFAD.
- IFAD and BMGF. (2014). *Impact assessment of the "Developing Business with the Rural Poor" programme in Běn Tre, Viet Nam (2008-2013) in Běn Tre, Vietnam*. Report of the pilot application of a participatory impact assessment and learning approach (PIALA). IFAD.

- IFAD, and BMGF. (2015). *Improved learning initiative for the design of a participatory impact assessment and learning approach (PIALA): Methodological reflections following the second PIALA pilot in Ghana*. IFAD.
- Kusek, J. Z., and Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. World Bank Publications.
- Levy, S., and Barahona, C. (2002). *How to generate statistics and influence policy using participatory methods in research* (Working Paper). Statistical Services Centre, University of Reading.
- Merrifield, J. (2002). *Learning citizenship* (IDS Working Paper No. 158). IDS.
- Mertens, D. (2009). *Transformative Research and Evaluation*. New York: Guilford Press.
- Ministry of Food and Agriculture (Government of Ghana) [MOFA/GOG], IFAD and BMGF. (2015). Final report on the participatory impact evaluation of the Root and Tuber Improvement and Marketing Program (RTIMP). Pilot application of a participatory impact assessment and learning approach (PIALA). IFAD.
- Mosse, D. (2001). 'People's knowledge', participation and patronage: operations and representations in rural development. In: Cook, B. and Kothari, U., (eds.), *Participation – the new tyranny?* London: Zed Press.
- Newman, D. (2015). *From the front of the room*. Rome.
- OECD. (2010). *DAC Guidelines and reference series quality standards for development evaluation*.
- Patton, M. Q. (2011). *Essentials of utilization-focused evaluation*. SAGE.
- Pawson, R. (2013). *The science of evaluation: a realist manifesto*. SAGE.
- Picciotto, R. (2012). *Experimentalism and development evaluation: Will the bubble burst?* Evaluation, 18(2).
- Punton, M., and Welle, K. (2015). *Straws-in-the-wind, hoops and smoking guns: What can process tracing offer to impact evaluation?* (10).
- Rihoux, B., and Ragin, C. C. (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. SAGE.
- Rogers, P. (2009). Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of Development Effectiveness*, 1(3).
- Rogers, P. and Peersman, G. (2014). *Developing a Research Agenda for Impact Evaluation in Development*. *IDS Bulletin* 45/6. Brighton: IDS.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. (Working Paper No. 38). DFID.
- Stern, E. (2015). *Impact evaluation: A guide for commissioners and managers*. Bond.
- Van Hemelrijck, A. (2013). Powerful beyond measure? Measuring complex systemic change in collaborative settings. In: *Sustainable Participation and Culture in Communication: Theory and Praxis*. Intellect Ltd.
- Van Hemelrijck, A. (2016a). *Governance in Myanmar. Evaluation using PIALA (participatory impact assessment and learning approach) of the "Building Resilient Livelihoods in the Dry Zone" project in the Dry Zone*. (Effectiveness Review Series 2015/2016.). Yangon: Oxfam GB. Available at: <http://policy-practice.oxfam.org.uk/publications/governance-in-myanmar-evaluation-of-the-building-equitable-and-resilient-liveli-620177>.

- Van Hemelrijck, A. (2016b). *Understanding “rigour” in participatory impact evaluation for transformational development: Insights from piloting a participatory impact assessment and learning approach (PIALA)* (IDS PhD Work in Progress paper.). Brighton: IDS.
- Van Hemelrijck, A. (2016c). *Stretching boundaries of evaluation practice with PIALA in Ghana*. (Blog post). Evaluation for Africa.
- Van Hemelrijck, A. (2017). *Walking the talk with participatory impact assessment and learning approach (PIALA) in Myanmar*. Oxfam GB Real Geek.
- Van Hemelrijck, A., and Guijt, I. (2016). *Balancing inclusiveness, rigour and feasibility. Insights from participatory impact evaluations in Ghana and Vietnam*. CDI Practice Paper, (14).
- White, H. (2014). Current challenges in impact evaluation. *European Journal of Development Research*, 26.
- White, H., and Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework* (Working Paper No. 15). International Initiative for Impact Evaluation (3ie).
- White, S. (2015). Qualitative perspectives on the impact evaluation of girls’ empowerment in Bangladesh. *Journal of Development Effectiveness*, 7(2).
- Winters, P., and Garbero, A. (2015, August). *Challenges to impact evaluation: Solutions for IFAD*. IFAD.
- Woolcock, M. (2009). Toward a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness*, 1(1).
- Woolcock, M. (2013). Using case studies to explore the external validity of “complex” development interventions. *Evaluation*, 19(3), 229–248.

The IFAD Research Series

01. **Agricultural and rural development reconsidered**
A guide to issues and debates
By Steve Wiggins
02. **Migration and transformative pathways**
A rural perspective
By David Suttie, Rosemary Vargas-Lundius
03. **Fostering inclusive outcomes in sub-Saharan African agriculture**
Improving agricultural productivity and expanding agribusiness opportunities
By David Suttie, Rui Benfica
04. **The effects of smallholder agricultural involvement on household food consumption and dietary diversity**
Evidence from Malawi
By Rui Benfica, Talip Kilic
05. **Rural-urban linkages and food systems in sub-Saharan Africa**
The rural dimension
By Karim Hussein, David Suttie
06. **Why food and nutrition security matters for inclusive structural and rural transformation**
By Steven Were Omamo
07. **Measuring IFAD's impact**
Background paper to the IFAD9 Impact Assessment Initiative
By Alessandra Garbero
08. **Fostering inclusive rural transformation in fragile states and situations**
By Karim Hussein
09. **Social protection and inclusive rural transformation**
By Carolina Trivelli, Silvana Vargas, Jhonatan Clausen
10. **Inclusive Finance and Inclusive Rural Transformation**
By Calum G. Turvey
11. **Food safety, trade, standards and the integration of smallholders into value chains**
A review of the literature
By John Humphrey
12. **An evidence-based assessment of IFAD's end of project reporting**
By Bia Carneiro, Alessandra Garbero
13. **Graduation models for rural financial inclusion**
By Khalid El Harizi, Xinjia Yan
14. **Disbursement performance of the International Fund for Agricultural Development (IFAD)**
An in-depth analysis of drivers and trends
By Tim Balint, Daniel Higgins, Paola Mallia, Silvana Scalzo, Paul Winters
15. **Remittances, growth and poverty reduction in Asia**
A critical review of the literature and new evidence from cross-country panel data
By Katsushi S. Imai, Bilal Malaeb, Fabrizio Bresciani
16. **Getting the most out of impact evaluation for learning, reporting and influence**
Insights from piloting a Participatory Impact Assessment and Learning Approach (PIALA) with IFAD
By Edward Heinemann, Adinda Van Hemelrijck, Irene Guijt



International Fund for Agricultural Development

Via Paolo di Dono, 44 - 00142 Rome, Italy

Tel: +39 06 54591 - Fax: +39 06 5043463

Email: ifad@ifad.org

www.ifad.org

www.ruralpovertyportal.org

 ifad-un.blogspot.com

 www.facebook.com/ifad

 [instagram.com/ifadnews](https://www.instagram.com/ifadnews)

 www.twitter.com/ifadnews

 www.youtube.com/user/ifadTV

ISBN 978-92-9072-767-5



9 789290 727675

