Department for International Development

iod parc

Evaluation, Quality Assurance, & Learning Service
EQUALS

LEARNING REVIEW OF RECENT AGRICULTURE EVALUATIONS

DFID EQUALS Learning Review Series

APRIL 2020

DFID Cover Note

This light touch learning review draws together lessons from recent evaluations of DFID-funded commercial agriculture programmes in Africa. It is part of DFID's work to strengthen learning from its programme portfolio across sectors and geographies and to make best use of the evidence generated by DFID-funded evaluation studies.

This review was commissioned by the Evaluation Unit for learning purposes and should not be seen as DFID policy. It is being published so that organisations working in this field can benefit from the findings and to contribute towards improved impact and effectiveness of donor-funded interventions supporting commercial agriculture.

The review was undertaken by two independent consultants, James Morton and Karin Tang, engaged through the DFID programme, Evaluation Quality Assurance and Learning Services (EQUALS). The findings of this review are the authors' own perspectives and do not necessarily represent the views of DFID.

Acknowledgements

The Evaluation Unit would like to thank the consultants, James Morton and Karin Tang, for their hard work and insights in this report and the EQUALS team (Jo Kaybryn and Karolina McLellan) for their support.

We would also like to thank all DFID colleagues who have supported this learning review through their feedback and comments, particularly Jackie Johnson, Watipatso Makandawire, Chris Mellor, Lola Russell, Siobhan Sutton, Frances Wood (Africa Regional Department); Simon Calvert and Liz Kirk (Growth and Resilience Department).

Louise Davis, Evaluation Advisor (Thematic Lead)
DFID Evaluation Unit
June 2020
louise-davis@dfid.gov.uk/Louise.Davis1@fcdo.gov.uk

List of acronyms

AAER	adopt-adapt-expand-respond
AECF	Africa Enterprise Challenge Fund
ARD	Africa Regional Department
CAPR	Commercial Agriculture Portfolio Review
CFA	Conceptual Framework on Agriculture
CMA	collateral management arrangement
CMO	context-mechanism-outcome
CSA	climate-smart agriculture
DAC	Development Assistance Committee
DFID	Department for International Development
ECF	Enterprise Challenge Fund
ELAN	DRC Private Sector Development Project
EQUALS	Evaluation Quality Assurance and Learning Services
FM	fund manager
FTESA	FoodTrade East and Southern Africa
GAP	good agricultural practice
ICT	information and communications technology
IMSAR	Improving Market Systems for Agriculture in Rwanda
LEAD	Livelihoods Enhancement through Agricultural Development
M4P	markets for the poor
MADE	Market Development in the Niger Delta
MEL	monitoring, evaluation and learning
MRM	monitoring and results management
MSC	market systems change
MSD	market systems development
MTE	mid-term evaluation
NAIC	net attributable income change
NGO	non-governmental organisation
OECD	Organisation for Economic Cooperation and Development
PMU	project management unit
PropCom	Rural and Agriculture Markets Development programme for Northern Nigeria
SRO	Senior Responsible Owner
TA	technical assistance
ToC	theory of change
ToR	terms of reference
VfM	value for money
WAFM	West Africa Food Markets

Contents

Executive Summary

DFID has commissioned this light-touch review to identify and synthesise learning around the design of commercial agriculture programmes as well as approaches to evaluating such programmes. The review focuses on a set of evaluation products published by DFID since April 2018 on eight programmes:

- FoodTrade East and Southern Africa (FTESA)
- Rural and Agriculture Markets Development programme for Northern Nigeria (PropCom)
- Market Development in the Niger Delta (MADE)
- West Africa Food Markets (WAFM)
- ELAN programme in DRC
- Livelihoods Enhancement through Agricultural Development (LEAD)
- Improving Market Systems for Agriculture in Rwanda (IMSAR
- Africa Enterprise Challenge Fund (AECF)

The review has addressed five key questions as follows.

1. What are the areas of common learning across the evaluations? For the ARD multi-country evaluations – do the evaluations identify any specific advantages or challenges involved with taking a regional or multi-country approach to provide support in this area?

Table 1: Summary of common learning

Lesson	Summary
The M4P and associated models are only partially demonstrated.	This is for two main reasons: i) because expected (usually higher-level) results have not materialised in the allocated timeframe and ii) because of weaknesses in the design including partial or selective use of available analysis and weak theory of change assumptions. In particular, there are trade-offs to be made between short-term results and systemic change. MADE and ELAN have performed well at the piloting stage, but have been unable to catalyse the 'respond' and 'expand' results that characterise the crowding-in phase of the adopt-adapt-expand-respond (AAER) process through which systemic change is expected to take place.
Behaviour change is challenging. Even more so without a good understanding of beneficiaries.	This tension between the pilot and crowding in stages of the AAER model highlights the step change in the challenge that programmes face in building momentum and critical mass to enable sustainable, systemic and behaviour change. Key to this is a clear and nuanced understanding of target beneficiaries. The FTESA and AECF MTEs highlight the need for rigorous beneficiary profiling to understand and design for the differentiated needs and preferences of the target beneficiary groups. Programmes do not appear to have sufficiently appreciated the strong incentives for existing livelihood strategies and the degree of intervention required to shift behaviour from 'hanging in' subsistence strategies.

Lesson	Summary
Deeper, more rigorous analysis is needed to strengthen programme results and impact.	Several of the evaluations report that delivery effectiveness and potential impact have been weakened by inadequate analysis and theory of change design at the outset. There is a need for honesty in the design about what can reasonably be expected to work, and at what scale results can be expected (with implications for DFID in terms of how it incentivises goal-setting). This is particularly the case in terms of understanding and adjusting for the external environment in which projects are being implemented; ELAN provides a good example of this, with its flexible, responsive approach. There is also a need to strengthen how ToC assumptions are identified and tested, particularly with respect to how expansion and response will occur.
Goals and targets need to be realistic and indicators need to be multi-dimensional.	Design/theory of change weaknesses have implications for the goals and targets being set for programmes, with an evident 'optimism bias' around what can be expected to be achieved. The potentially distorting effect of targets is reflected in those evaluations which have commented on the measurement of beneficiary impact, noting that it is not enough to just count the beneficiaries – indicators need to demonstrate to what extent beneficiaries benefit (depth of impact). This suggests the need for multi-dimensional indicators for measuring beneficiary impact.
Get the timeframe and timing right: establish readiness.	A key constraining factor identified by a number of evaluations is the timeframe within which results are expected to be delivered. This suggests the need to match expectations to the timeframe – where it is not feasible to plan for longer programmes. Rolling out too quickly can lead to mistakes and insufficient time to 'learn by doing'.
Accept that there will be a trade-off between results and reach.	A common theme has been the absence of evidence on pro-poor impact. Clearly there are trade-offs to be made between viability (at the programme as well as project levels) and real poverty reduction impact. The ELAN evaluation concludes that, 'In the short-term at least, M4P interventions are unlikely to be an effective way to reach the very poorest and most vulnerable'.
Cross-programme coordination and partnerships can strengthen results and impact.	The multi-factoral nature of the M4P approach means that there are significant opportunities to benefit from synergies within and across programmes through cross-programme coordination and partnerships. Different projects may address different dimensions of the same problem (PropCom); or one project may provide an entrance for the products or services of other projects (AECF).
Test and iterate to maximise results.	Programmes need to re-evaluate on a regular basis, and adjust accordingly. The IMSAR review of the intervention design process promotes a 'learning by doing' ('learning by failing') approach in which 'pragmatically derived' interventions are implemented (and presumably allowed to fail in some cases) for their learning potential rather than for delivery of results.
Programmes struggle to address difficult tasks and mandates – gender and VfM.	Cross-cutting issues of gender and value for money were in general weakly addressed across the programmes evaluated. Evaluations themselves demonstrated limited understanding of the complexity of gender patterns that need to be reflected in the design of interventions as well as evaluation approaches; and while most report on programme approaches to measuring and monitoring VfM, they generally conclude these were unsatisfactory.
Advantages and challenges of regional or multi-country approaches.	The evaluations of the two regional programmes do not dwell on advantages and challenges of regional or multi-country approaches vis-à-vis single country approaches; and no lessons stand out from this learning review that are specific to the regional coverage of the FTESA and WAFM programmes. This highlights a potential area for future research.

2. To what extent are the outcome and impact statements of the programmes examined coherent with the directions set out in DFID's Conceptual Framework on Agriculture?

There is a high degree of consistency across the programmes with the CFA (which is largely a function of the common use of the M4P approach):

- At impact level, six programmes target poverty reduction and livelihoods, while the two regional programmes target food security.
- Outcome-level statements focus on better functioning, more accessible markets – with the exception of LEAD, which focuses on smallholder incomes.

All programmes are clustered within the agricultural transformation and 'stepping up' strategy, underpinned by the focus of M4P and market systems interventions on value chain development and greater participation by poorer farmers.

At the level of log frame impact and outcome statements, therefore, the set of programmes is consistent with the CFA. However, this is largely by virtue of phrasing and the general level at which these statements are made (see Annex 2). It does not necessarily follow that the programmes are more meaningfully coherent with the CFA at lower levels of the logic model, in terms of their theory of change assumptions and the implications for how programmes are designed.

3. What were the advantages/disadvantages of the evaluation methods used? Would other forms of evidence generation provide more effective ways to meet evidence gaps?

The set of evaluations reviewed includes a mix of mid-term and final evaluations, some of these conducted as part of a longitudinal evaluation process; plus one interim review and the two IMSAR topic-focused reviews. These adopt formative or summative approaches depending on the stage in the programme, and apply a generally standard set of evaluation methodologies. Table 5 provides a summary, based on the intentions stated in the methodology section of the evaluation reports.

Evaluations cannot depend on availability or quality of programme data. Dependence by the evaluation on partner reporting – and lack of confidence in reported results – has been an overriding theme. Data (commonly VfM data) has been absent, has been incomplete or of low quality. The WAFM evaluation highlights **underlying causes of resource availability and grantee capacity vis-à-vis the reporting burden**. A particular challenge has been to convincingly report against high-level indicators for beneficiary numbers and income impact.

Methodological approaches are broadly standard across the set of evaluations. Most apply theory-based methods, taking a mixed methods approach, gathering qualitative and quantitative data through document review, interviews and focus group discussions – with some additionally including some form of survey – and almost all report against OECD DAC evaluation criteria. The use of a **realist enquiry approach**, in combination with **contribution analysis**, and a **baseline case evaluation** to establish context-mechanism-outcome (CMO) configurations, helped the FTESA evaluation to convincingly organise and synthesise a large volume of evidence against the evaluation questions, but the WAFM evaluation, using the same methodology, highlights that this a very time-consuming exercise.

Four evaluations use quantitative survey data to help establish (attribute) the role of the intervention in bringing about the observed change. None of the evaluations, however, gives great prominence to survey results. This may reflect unarticulated concerns about the reliability of the data or its interpretation; or the fact that findings from only a few of several grants cannot easily be extrapolated to other grants and other markets. What is also surprising is the general absence of counterfactual analysis and assessment of attribution, given the investment in including control groups. If survey results are so inconclusive in a sector where it is relatively easy to count things, **surveys and impact evaluations need to be more carefully designed to ensure they are able to answer evaluation questions with a sufficient level of confidence to ensure VfM**.

4. Where the evaluations examined common questions or theory of change (ToC) assumptions, to what extent do they provide consistent evidence for or against these?

We collated the evaluation questions from across the seven programmes evaluated to draw out the most common areas of inquiry (paraphrased in Table 7).

Table 2: Summary of common questions and findings

Question	Findings
Is the programme delivering VfM?	There is a general dissatisfaction with the quality and coverage of VfM reported data and analysis, and evaluations are consequently reluctant to make definitive conclusions on VfM, or conclusions that are not heavily qualified. A number of common indicators rely on available and relatively verifiable data – but what they actually measure varies so considerably that any comparison or benchmarking becomes difficult.
To what extent is the programme contributing to or likely to contribute to systemic change?	Those evaluations that posed this question comment that it is either too early, or the duration of the programme is too short, to be able to confidently assess the likelihood for systemic change.
How effective has the programme been in delivering log frame outcomes and impacts?	The PropCom and MADE evaluations are generally positive on log frame performance, the ELAN evaluation less so.
To what extent does programme governance and management underpin delivery?	This is an important question for VfM, and a concern across all evaluations. The four evaluations that had it as a formal evaluation question considered the programme to have 'room for improvement', for both internal and external reasons. ELAN, however, stands out as having driven positive results through flexible and responsive management.
How well is gender integrated into the programme?	All programmes were reporting gender disaggregated data to some extent, and were able to show female participation, albeit below target in some cases. However, there is limited accompanying analysis.
To what extent has improved availability and use of inputs helped deliver expected results?	This is a specific question for the FTESA, WAFM and LEAD evaluations, with positive – but qualified – results reported for each of them.
Does the programme have robust results monitoring systems in place?	This is a question for ELAN, LEAD and AECF, with a number of reservations articulated – also reflecting the dependence of the evaluation on programme-generated monitoring data.
Are results relevant to beneficiary needs, including marginalised groups?	This is an evaluation question for PropCom, WAFM and LEAD, with findings reflecting the rather inconclusive findings noted in Section 1 on the trade-offs between results and reach.

5. Were there any opportunities to build in practical measures to make it easier to synthesise learning across programmes in future? For example, are there any common evaluation questions, or common metrics, or common value for money (VfM) measures, that have been used successfully across more than one of the evaluations reviewed?

In theory, there should be good scope for synthesis based on the number of common evaluation questions, the commonality of certain key issues and the applicability of standard, theory-based evaluation methods. In practice, this is limited by the inability to be confident in programme-generated monitoring data, and incomplete data collection; insufficient comparability across ‘comparators’; and the large and often unaddressed role played by context and externalities. This is compounded by inconclusive findings from gender and VfM assessments.

Based on the findings emerging from the review, we have identified three avenues that hold potential for the purposes of synthesis:

1. Approaches for further exploration

The FTESA evaluation used two methods that might be further explored for their potential for standardisation across other evaluations: use of the **AAER framework** to assess systemic change and use of a **baseline case evaluation**. It would be useful to be able to review the AAER analytical framework applied (to the extent that it has been tailored from the Springfield model) and consider its applicability across a longer duration where the ‘expand’ and ‘respond’ results might be expected. It could also be fruitful to review the analytical framework for the baseline case evaluation, and to review the baseline and final case evaluations side-by-side to understand how ToC linkages have been traced and how this has surfaced evaluation conclusions.

2. Use of common metrics

It is not very practical to try to aggregate micro-level results (such as NAIC) where small differences in calculation methods can be amplified into large margins of error at the aggregate level. Further, the review has noted the common challenge of imposing data collection and reporting burdens on implementing partners with limited resources, capacity or will to do this well enough to be reliable. An alternative would be to start with an assessment of how much data, at what level of granularity, is ‘good enough’ to be able to confidently answer high-level questions about whether a programme is making a positive difference, and whether this is enough vis-à-vis the level of investment (i.e. value for money). This calls for simple models that are robust to data variability within a relatively large margin of error. Two potentially ‘good enough’ methods might be the use of **evaluative rubrics**[1] and **multi-dimensional indicators**. Both aggregate input data to a higher generic level, allowing for comparability across diverse types of data and can smooth out some of the volatility of individual datasets.

3. Development and guidance on gender and VfM approaches

It might be practical for DFID to provide a stronger steer and guidance on what it expects in these two areas from evaluations of M4P programmes, and this might require some preliminary analysis.

There is a substantial literature on the dynamics of women’s participation and role across the three livelihood strategies and within an M4P context. This can be used to frame a common analytical approach tailored to the market systems framework, on the basis of which specific, relevant questions about gender can be articulated. This common approach would then provide a foundation for synthesis across different evaluations.

In the case of VfM, while guidance and a broad ‘4Es’ framework exist, this has not resulted in consistent approaches across the evaluations that might support synthesis. This may be because evaluators have been too narrow in their interpretation of the 4Es framework. The 4Es do not represent all of the possible dimensions of VfM, as recent thinking by Julian King illustrates[2]. A VfM framework, based on the 4Es approach but more tailored to market systems programmes, might support better evidence and stronger conclusions.

[1] https://www.betterevaluation.org/en/evaluation-options/rubrics

[2] https://www.julianking.co.nz/vfi/criteria/

Introduction

DFID has commissioned this light-touch learning review to identify and synthesise learning around the design of commercial agriculture programmes as well as approaches to evaluating such programmes. The review focuses on a set of evaluations published by DFID since April 2018 which provides an opportunity to draw together areas of common learning, as well as to consider how such evaluations might be adapted in future to enable more rigorous synthesis of findings and better learning on what works to reduce poverty through DFID agricultural programming. In Sections 1-5 we set out the main review findings against each of the five questions in turn.

Context and background

Since February 2019, DFID's Evaluation Unit has been developing a new centrally managed programme to conduct thematic evaluations and mechanisms to support greater coordination of evaluations at a thematic level, within priority sectors and/or geographies. It is also exploring the potential to promote greater use of consistent and comparable outcome and impact indicators within thematic areas as a means to draw together evidence and learning. In parallel, DFID's Growth and Resilience Department plans to develop a set of voluntary core indicators for new DFID commercial agriculture programmes, following recommendations made by the second **Commercial Agriculture Portfolio Review** (CAPR) published in 2019, which focuses on coherence of programming with shifts set out in DFID's **Conceptual Framework on Agriculture** (CFA) (2015), collation and verification targets, results and budgets and analysis of specific aspects of results data..

The learning review is intended to be complementary to the portfolio review, by drawing out learning on evaluation methods and indicators used across more than one intervention.

Review purpose, questions and scope

The purpose of the review is:

- To draw together areas of common learning across these evaluations in relation to relevance to DFID policy and programme effectiveness, with a view to improving programme design and implementation.
- To inform thinking on how to improve capability to evaluate impact and cost-effectiveness, and DFID's contribution, through synthesis of learning from interventions with similar objectives and activities.

Review questions

The terms of reference (ToR) identified five questions as the focus for the learning review:

1. What are the areas of common learning across the evaluations?
    - For the ARD multi-country evaluations – do the evaluations identify any specific advantages or challenges involved with taking a regional or multi-country approach to provide support in this area?
2. To what extent are the outcome and impact statements of the programmes examined coherent with the directions set out in DFID's Conceptual Framework on Agriculture?
3. What were the advantages/disadvantages of the evaluation methods used? Would other forms of evidence generation provide more effective ways to meet evidence gaps?
4. Where the evaluations examined common questions or theory of change assumptions, to what extent do they provide consistent evidence for or against these?
5. Were there any opportunities to build in practical measures to make it easier to synthesise learning across programmes in future?

- For example, are there any common evaluation questions, or common metrics, or common value for money (VfM) measures, that have been used successfully across more than one of the evaluations reviewed?

Scope

The review has taken six agriculture programmes as its main focus; in addition to one other programme that has applied alternative learning approaches (IMSAR) and a further multi-donor funded programme (AECF). All evaluations were reviewed by DFID's Evaluation Quality Assurance and Learning Services (EQuALS) and received at least a 'satisfactory' rating.[3] Table 3 provides an overview.

Table 3: Programmes and evaluations under review

Programme	Department	Period	Type of evaluation & publication date
Focus programmes			
FoodTrade East and Southern Africa (FTESA)	Africa Regional Department	2013-18	Mid-term (2017) and final (2018) evaluations
Rural and Agriculture Markets Development programme for Northern Nigeria (PropCom)	DFID Nigeria	2013-20	Final evaluation (2018)
Market Development in the Niger Delta (MADE)	DFID Nigeria	2014-20	Independent review (interim report 2019)
West Africa Food Markets (WAFM)	Africa Regional Department	2013-19	Mid-term evaluation (2018)
ELAN programme (agriculture-related components)	DFID DRC	2012-24	Mid-term evaluation (2018)
Livelihoods Enhancement through Agricultural Development (LEAD)	DFID Tanzania	2013-17	Mid-term evaluation (2016)
Programmes with alternative learning approaches			
Improving Market Systems for Agriculture in Rwanda (IMSAR)	DFID Rwanda	2015-22	Two quarterly topic-focused reviews[4] (2019)
Multi-donor commercial agriculture evaluations and reviews			
Africa Enterprise Challenge Fund (AECF)			Mid-term evaluation (2015)

This covers a range of programmes at varying stages of implementation and of differing durations, with the oldest and longest (ELAN) starting as early as 2012 and continuing until 2024. Three are completed and another two due to complete in 2020. The set of evaluations is equally diverse in terms of evaluation type and stage, scope and depth of analysis. The MADE interim review is relatively light touch, and the IMSAR topical reviews narrow in scope; while the mid-term evaluations (MTEs) tend towards formative evaluation rather than reporting on results as is more usual at the final evaluation stage. This is reflected in the quantity and quality of evidence emerging from across the programme portfolio against the review questions.

[3] The MADE Independent Review was not quality assured by EQuALS as this is only mandatory for DFID evaluations.

[4] The contract originally provided for four quarterly MEL outputs per year and has subsequently been revised to two. For the purposes of this review, two reviews were made available from 2019.

Methodology and process

The learning review has been carried out in two stages. In the first stage, an agriculture expert carried out a desk review of relevant documents, and a first draft review report was completed in November 2019.

The first draft was reviewed by DFID and, in discussion with EQUALS, a restructured second draft was prepared by an editor, to focus more strongly on extracting the lessons and improving the usability of the first draft findings. This second draft has, in turn, been reviewed by DFID and the current report is the revised report incorporating DFID's feedback.

Limitations

Key limitations have included a limited timeframe and time allocation and the entirely desk-based nature of the review, with relevant staff and programme managers largely unavailable for interview. The review is based primarily on review of the completed evaluation reports, with some reference to business cases, log frames, annual review reports and programme completion review reports. Thus our understanding of operating contexts and what happened during implementation of the programmes under consideration is based predominantly on the evidence and conclusions presented in the evaluation reports.

It was found on closer analysis that not all the programmes were suitable for inclusion: it was agreed early on that Enabling the Business of Agriculture should be left out as there was limited overlap with the other evaluations included. Three further programmes were omitted (Strengthening Indian Trade and Investment for Africa, Northern Uganda: Transforming the Economy through Climate Smart Agribusiness and Global Agriculture and Food Security Programme) due to lack of detail in the evaluation, lack of overlap, or taking a different approach.

Overview of programme portfolio

Table 4 summarises the key objectives and activities of the programmes under review.

Table 4: Summary of programme objectives and activities

Programme	Summary
FTESA FoodTrade East and Southern Africa	FTESA aimed to catalyse lasting changes that enable efficient trade in staple foods across the region to improve the lives of farmers, suppliers, service providers, traders, retailers and consumers, contributing to price and market stability for staple foods in the region. FTESA aimed to invest in systems that allow small-scale farmers to access regional grain markets. It supported activities across three output areas: 1. Improved post-harvest markets (storage and aggregation; market information; value chain coordination; warehouse receipts and supplier credit; grades and standards) 2. Improved input markets (including seeds and fertilisers) 3. Improved trade environment and reduced uncertainty
PropCom Rural and Agriculture Markets Development programme for Northern Nigeria	PropCom, in its second phase, aims to increase employment and improve productivity in selected rural and agricultural market systems in northern Nigeria. Its outputs focus on: 1. Supporting rural and agricultural market systems to work more effectively for poor farmers and small-scale rural entrepreneurs; 2. Promoting the adoption of climate smart agriculture (CSA) practices; 3. Influencing private investors, government, non-government organisations and development agencies to make changes in their approach to northern Nigeria;

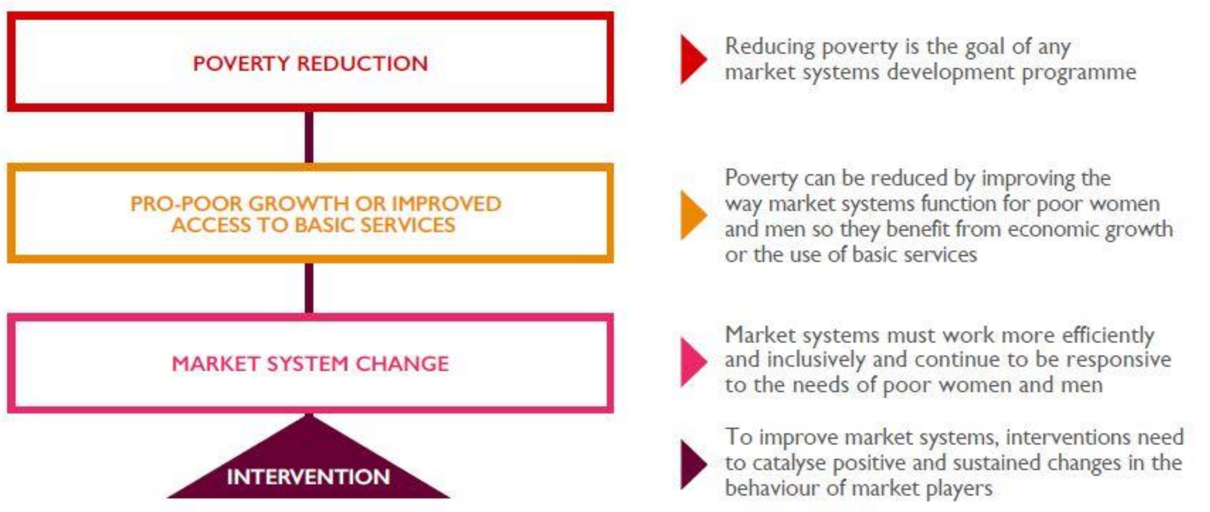
Programme	Summary
	4. Implementing high-risk interventions that have the potential to bring about significant change in market systems.
MADE Market Development in the Niger Delta	MADE uses a market systems approach to generate wealth creation and employment in the Niger Delta's non-oil economy, supporting palm oil, poultry, fisheries, cassava, and agricultural inputs and expanding into ICT, hospitality and creative industries in its second phase. It aims to facilitate pro-poor growth in these markets by: 1. Introducing new and/or improved inputs, products, services and technologies that benefit poor people in its target markets; 2. Influencing development agencies, support service providers (private, public, and NGO) and private investors to change their approach to engaging with the poor in the Niger Delta region.
WAFM West Africa Food Markets	WAFM aims to tackle multiple failures in staple food markets in the region, stimulating staple food trade along the key corridors across two borders: between Ghana and Burkina Faso and between Niger and Nigeria and increasing income levels and food security for producers and consumers. It aims to do this by: 1. Establishing a Policy Facility to generate evidence, identify and implement policy interventions; 2. Developing improved business models for the production and marketing of staple foods across the border; 3. Delivering better business models for marketing of staple foods.
ELAN ELAN programme (agriculture-related components), Private Sector Development Programme	ELAN is a flexible facility with market systems and sectors identified to address a variety of market, government, information and coordination failures in six key economic sectors associated with high pro-poor growth potential. In agriculture these include coffee, cocoa, maize and rice. It works through pilots in partnership with private sector firms providing market insight, technical assistance and loan guarantee funds to address market constraints.
LEAD Livelihoods Enhancement through Agricultural Development	LEAD aims to increase the income of smallholder maize and poultry farmers in Tanzania through support across four output areas: 1. Access to local, regional and national markets; 2. Improved quality and availability of inputs and technologies for smallholder farmers: 3. Improved access to agrifinance by smallholder farmers: 4. Addressing demonstrable gaps in the value chain for maize and poultry.
IMSAR Improving Market Systems for Agriculture in Rwanda	IMSAR aims to catalyse agricultural commercialisation in Rwanda by identifying market failures – such as limited access to finance and technology, limited or poor storage, lack of access to inputs such as fertilisers – and developing initiatives to address them; by strengthening the capacity of farmers and private sector actors to engage with, and benefit from, markets; and by providing short-term and long-term investment for small and medium size agriculture businesses – where possible, partnering with existing specialist financial institutions.
AECF Africa Enterprise Challenge Fund	AECF is an autonomous, multi-donor funded challenge fund that aims to catalyse private sector entrepreneurs in Africa to innovate and find profitable ways of improving market access and functioning for the poor. It targets agribusiness and access to finance. This is to be achieved by:

Programme	Summary
	1. Identifying, selecting, contracting and providing managed support to business initiatives; 2. Identifying and supporting business initiatives which generate sustained and significant positive development impacts and which are innovative; 3. Supporting business initiatives (and the products and services delivered through them) to be sufficiently commercially viable that associated development impacts are sustainable; 4. Supporting business initiatives to generate development benefits with high outreach, in particular to: those in rural areas; very poor; women; and youth.

Theoretical underpinnings

An overriding theme is the **markets for the poor** (M4P) approach which aims to tackle market failures and strengthen the private sector so as to enable greater economic participation and improve the lives of the poor. More recently, this has evolved into a **market systems development** (MSD) or **market systems change** (MSC) approach which focuses on improving the poor's terms of participation within systems for economic exchange and basic service delivery.

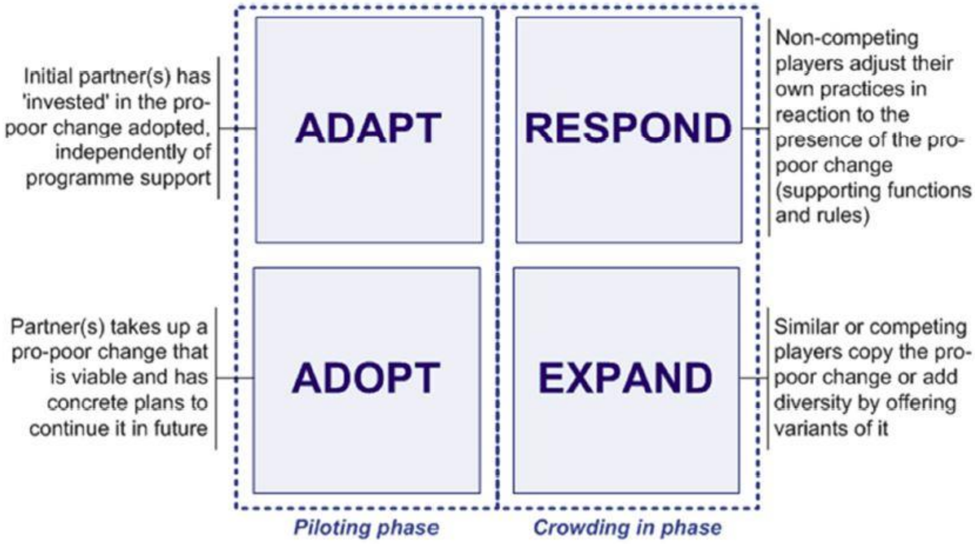
Figure 1: Strategic framework for market systems development



Source: The Operational Guide for the Making Markets Work for the Poor (M4P) Approach, SDC, 2015

Market systems approaches generally encompass an adopt-adapt-expand-respond (AAER) framework to help with the management and measurement of systemic change. The AAER framework provides an analytical lens for understanding the process through which programmes expect that work with selected partners piloting innovations in target markets will promote wider change by 'crowding in' as other market actors adopt those innovations for themselves (Figure 2).

Figure 2: Systemic change framework



Source: Adopt-Adapt-Expand-Respond: a framework for managing and measuring systemic change processes, The Springfield Centre, 2014

Of the eight programmes under review, five have followed the M4P approach in individual countries (PropCom, MADE, ELAN, LEAD, IMSAR), while FTESA and WAFM have the additional aim of promoting the development of regional trade in food staples, while working along similar M4P lines.

1. Common learning across the programme portfolio

Based on a mapping of the conclusions and lessons identified by the set of evaluations and reviews, the following areas of common learning have been highlighted. Many of these lessons are unsurprising, and common to evaluations in general, across diverse sectors; but the analysis draws out rich theoretical and programmatic insights that are specific to agriculture and M4P approaches.

a. The M4P and associated models are only partially demonstrated

Across the programmes under evaluation, the M4P model is only partially demonstrated, for two main reasons: i) because expected (usually higher-level) results have not materialised within the allocated timeframe and ii) because of weaknesses in the design including partial or selective use of available analysis and weak theory of change assumptions (see also b and c below).

Successes are reported in pockets by ELAN (coffee), LEAD (poultry) and provisionally (at interim stage) for MADE. Single sector success is attributed to a broad sector approach at an early stage in the ELAN evaluation which reports that 'Significant early MSCs have proved the relevance, feasibility and impact of out-grower schemes and supply partnerships between dispersed smallholder growers, globally integrated exporters and international buyers … Significant MSC is thus evident in the coffee sector, which adopted a broader, sector-level approach much earlier in the implementation process than most other sectors.'

Tension between short-term results and systemic change. Two programmes, MADE and ELAN, have performed well at the piloting stage of the programme, but have been unable to catalyse the 'respond' and 'expand' results that characterise the crowding-in phase of the AAER process. In the case of ELAN, its success in achieving adoption and adaptation has generated 'impressive impact-level results' – but these have fallen short of systemic results. ELAN's success is attributed to a highly adaptive, opportunistic approach that prioritises rapid delivery of impact-level results. The importance of a flexible approach is emphasised by the LEAD evaluation, which notes that 'lack of facility within the programme to be flexible and responsive to market needs … limits the applicability of an M4P approach'. However, the IMSAR review of the market selection process cautions against allowing a focus on 'quick wins' to obscure objectives in terms of impact, opportunity and replication. The IMSAR review notes that a focus on rapid results through direct interventions risks diverting efforts away from catalysing systemic change and may not represent the most effective use of programme resources in the long term.

b. Behaviour change is challenging. Even more so without a good understanding of beneficiaries

This tension between the pilot and crowding in stages of the AAER model highlights the step change in the challenge that programmes face in building sufficient momentum and critical mass to enable sustainable, systemic and behaviour change. Key to this is a clear and nuanced understanding of target beneficiaries. Two evaluations – the FTESA and AECF MTEs – highlight the need for **rigorous beneficiary profiling to understand and design for the differentiated needs and preferences** of the target beneficiary groups. Several instances are recounted of target beneficiaries not responding to interventions in expected ways (i.e. ToC assumptions not holding): continued side-selling by farmers undermined FTESA's ambitions, price incentives were too small to motivate paravets under PropCom and both FTESA and AECF had high levels of defaults on loans. As the FTESA final evaluation points out, where farmers did not change behaviour, 'such behaviour reflects rational risk perceptions on the part of farmers unwilling to change practices, or maintain new practices, if expected benefits do not materialise'. This suggests that programmes have not sufficiently appreciated the strong incentives for existing livelihood strategies and the degree of intervention required to shift behaviour from 'hanging in' subsistence strategies (see Section 2).

The FTESA evaluations place significant emphasis on the need to build transparency and trust, to underpin 'proof of concept' and demonstration effects – and this takes time and resources to 'trickle through'. The PropCom evaluation provides a number of conclusions on what needs to be achieved to nudge household behaviour in the desired direction, including fast return on investment and low barriers to entry. The LEAD evaluation highlights the added value perceived by farmers in working as a group (e.g. through savings schemes) and the role of increased social cohesion.

c. Deeper, more rigorous analysis is needed to strengthen programme results and impact

Several of the evaluations report that delivery effectiveness and potential impact have been weakened by inadequate analysis and theory of change design at the outset. The ELAN evaluation highlights 'the need for an improved and operational concept of MSC which focuses on direct and comparable measures of improvements in market performance and terms of access for the poor' and a weak approach which assumes 'that wider impact beyond the pilots … can plausibly take place through market mechanisms alone, beyond the lifetime of the pilot'. Its diagnostic analysis is of 'variable quality' and at times missing altogether. The evaluation states, 'This has compromised the relevance and effectiveness of certain interventions … It has also contributed to poor MSC and impact outcomes where these have manifested'. The LEAD evaluation notes the absence of any analysis of the underlying causes of market failures; and the AECF evaluation highlights the weak commercial viability of a number of projects, raising questions about the rigour of the business planning process and 'whether the business plans as approved by the IC provide a realistic basis for selection and planning'. This points to the need for **honesty in the design** about what can reasonably be expected to work, and at what scale results can be expected (with implications for DFID in terms of how it incentivises goal-setting).

This is particularly the case in terms of understanding and **adjusting for the external environment** in which projects are being implemented. The successes achieved by ELAN – with its flexible, responsive approach – came about in spite of a weak business environment. The PropCom evaluation highlights the need to factor in external conditions when selecting potential intervention markets: 'this may mean changing focus on markets that benefit from a more favourable external environment'. Other programmes appear to discount the role of external factors over which they have no control: the AECF evaluation notes the systemic factors, including weak business environments that have negatively affected performance while the PropCom evaluation notes that external factors can have positive as well as negative effects.

Where analysis and design have been instrumental in driving results, this has been attributed to 'designing interventions that were relevant not only to market constraints but also to beneficiaries' needs' in the case of PropCom (with some interventions leading to copying of behaviour); while IMSAR recommends 'matching of interventions to root causes of market failure'.

There is a need to strengthen how ToC assumptions are identified and tested. Weak or inadequate analysis inevitably affects the quality of programmes' logic models. ELAN's ToC 'has insufficiently and incompletely articulated the assumptions … that are required for its results chain to hold, particularly with respect to how expansion and response will occur (after adoption and adaptation). There has also been insufficient articulation and testing of key assumptions in sector and intervention results chains, which may have contributed, in some cases, to inappropriate pilot and partner selection.' Similarly for WAFM, the logical foundations – the theory of change, its assumptions and the links between assumptions and interventions – were 'not as strong as they could have been'. The evaluation goes on: 'Using [existing] resources would allow the designers of programmes (and WAFM in its implementation) to avoid resourcing duplicative work and would also assist in interrogating the programme logic and clarifying its primary purpose'.

d. Goals and targets need to be realistic and indicators need to be multi-dimensional

The weakness of ToCs and ToC assumptions relying on partial or insufficient analysis has been commented on above. This has implications for the goals and targets being set for programmes, with an evident 'optimism bias' around what can be expected to be achieved. The AECF evaluation reports that (project-level) projections for development impact are often overrated and the FTESA MTE highlights the need to establish realistic ambitions; while the ELAN evaluation notes the potentially distorting effect of comprehensive quantitative targets at impact level that may have contributed to an over-emphasis on short-term achievement of measured impact results, rather than on sustainably strengthening market systems.

The potentially distorting effect of targets is reflected in those evaluations (FTESA and AECF MTEs) which have commented on the measurement of beneficiary impact, noting that it is not enough to just count the beneficiaries – indicators need to demonstrate to what extent beneficiaries benefit (depth of impact). This suggests the need for multi-dimensional indicators for measuring beneficiary impact. It applies also at output level: for instance, beneficiary engagement should go beyond numbers reached – there is, for example, a big difference between attendance at a one day demonstration event with 100 farmers and a five-day course in a group of 20.

e. Get the timeframe and timing right: establish readiness

A key constraining factor identified by a number of evaluations is the timeframe within which results are expected to be delivered. This suggests the need to **match expectations to the timeframe** – where it is not feasible to plan for longer programmes. The FTESA MTE comments on the challenge of attempting to achieve market-level changes through a five-year programme while the final evaluation reports that some grantees did not provide support for long enough to deliver the 'critical mass' of consistent success required to build buy-in for the intervention and encourage others to crowd in. The MADE evaluation notes that the duration of the programme has limited its ability to show larger results, particularly from a sustainability and systemic point of view. And the WAFM evaluation comments that influencing policy change takes time. PropCom notes that, 'Two of the programme's most successful interventions in terms of reach and efficiency were legacy interventions, indicating that M4P programmes take time to achieve scale.'

The negative impact of a compressed timeframe is illustrated by the FTESA where the final evaluation notes that 'some of the projects designed to deliver systemic change over a relatively short period of time attempted to reach scale too quickly, before positive demonstration effects emerged' and concludes that 'Rolling out too quickly can lead to mistakes and insufficient time to learn by doing and, more importantly, disincentivises participants'.

f. Accept that there will be a trade-off between results and reach

A common theme has been the absence of evidence on pro-poor impact. The PropCom, ELAN and AECF evaluations each explicitly state that evidence of pro-poor impact is limited and that in several cases the poorest have not benefited. In the case of ELAN, poor poverty targeting has played a clear role: 'Using a poverty threshold that includes 80% of the population and much of the middle class has militated against an effective focus of interventions on a clear target group of the 'poor''. This illustrates design decisions that make the achievement of results easier, at the expense of reach and the 'leave no one behind' agenda. Clearly there are trade-offs to be made between viability (at the programme as well as project levels) and real poverty reduction impact. The AECF evaluation notes that 'companies had changed their business model to focus on more commercial farmers as the focus on smallholders was not experienced as a profitable business opportunity by the grantees'. Similarly, the ELAN evaluation notes that some of the market opportunities facilitated by the programme may exclude low income producers who are not capable of responding; and raises concerns about the viability of collateral management arrangements (CMAs) and hybrid seed and out-grower scheme pilots which are targeting low-income smallholders, suggesting that, 'These weaknesses could be largely resolved by targeting emerging commercial farmer smallholders'. It concludes that, 'In the short-term at least, **M4P interventions are unlikely to be an effective way to reach the very poorest and most vulnerable**. Those best-placed to benefit are those with some level of human capital and other resources, who can take advantage of market opportunities created, or as consumers, those with purchasing power and proximity to markets'.

The overall impression is that programmes did not have the resources or the remit to develop the socio-economic profile needed to distinguish the capacities and needs of different classes in the target population; or to devise separate strategies to reach each class. Further, they did not confront difficult choices to be made in terms of poverty targeting: it may be that enabling the less poor to move forward will be the most effective way to support development, or that transfer payments of some kind are the only help that the poorest of the poor can use, at least in the short term. Programme designs appear to have avoided these difficult decisions, in the absence of a deeper analysis and understanding of rural society and rural livelihood strategies.

g. Cross-programme coordination and partnerships can strengthen results and impact

The **multi-factoral nature of the M4P approach means that there are significant opportunities to benefit from synergies** within and between programmes, through cross-programme coordination and partnerships: the PropCom evaluation notes, 'Very few market constraints are as a result of a simple oversight or missing market component. In many cases there are numerous factors affecting the lack of market activity. As a result it is often unlikely that a single programme partner can overcome these constraints. In at least two of the Propcom Mar-karfi interventions, (Poultry Health and EWRS), a number of partners played specific and defined roles in designing a successful business model'. Similarly, the AECF evaluation reports that 'several examples were found of cooperation and or synergy between different projects. Cases in which one project provided an entrance for the products or services of other projects, increasing the development impact of these projects'. This worked less well under the FTESA, where the design, 'which largely relied on promising interventions and interactions surfacing from grantees … reduced the ability to directly build in strong interlinkages and complementarities across the portfolio'. The FTESA MTE recommends that 'future portfolio-approach programmes are designed with sufficient resources and mechanisms up-front that ensure that PMUs (or similar) foster coordination and collaboration between implementing partners/grantees from the start'.

h. Test and iterate to maximise results

Getting the analysis and design right, as discussed above, is only part of the story. With constantly evolving external and internal contexts, there is a need for programmes to **re-evaluate on a regular basis, and adjust accordingly**. The ELAN evaluation highlights the need for the programme to test assumptions in design and implementation; and the IMSAR review of the intervention design process promotes a 'learning by doing' ('learning by failing') approach in which 'pragmatically derived' interventions are implemented (and presumably allowed to fail in some cases) for their learning potential rather than for delivery of results: 'Without that experience, both the identification of new opportunities and the design of new interventions would be less efficient.' The AECF evaluation also notes that a certain failure rate is inevitable and that the AECF Committee should agree on an acceptable rate – with an implicit understanding that this is built into a risk/return calculation designed to maximise impact.

i. Programmes struggle to address difficult tasks and mandates – gender and VfM

Cross-cutting issues of gender and value for money were in general weakly addressed across the programmes evaluated. Two programmes stand out on gender. ELAN 'has made concerted efforts to acknowledge and address women's structural position in the contexts of both the household and the market' with 'strong and sustainable gender-specific impact' and 'significant women's economic empowerment benefits'. This is in part attributed to impressive cross-cutting workstreams of gender and political economy leading to genuinely innovative practice. MADE achieved high female participation rates which is attributed to the selection of markets which have a significant presence of women. It included gender targets and targeted participation of and access by women. Across the other programmes, the FTESA's 'lack of a sharp gendered lens means that access for smallholders (especially female smallholders) will be under the terms of local gender norms, possibly further entrenching them'; PropCom has 'made little progress in increasing women's direct access and involving women more directly in the economy' – and the progress it did make was 'largely by default rather than design'; and WAFM lacked VfM analysis to better understand equity issues and the costs of reaching women farmers.

The evaluation of PropCom reports that 'Programme partners were not convinced of a business rationale for targeting women possibly because of overriding cultural norms regarding gender roles'. It notes that interventions that targeted women exclusively were more costly in terms of reach and less efficient in terms of generating both individual ROI and programme benefit. This increased cost of women specific interventions impacted on the probable sustainability and scale of the interventions, affecting partner participation. In both PropCom and IMSAR gender has been relatively neglected compared with the expectations established in the business case.

The evaluations themselves demonstrated limited understanding of the complexity of gender patterns that need to be reflected in the design of interventions as well as evaluation approaches. For example, women farmers may be single parents supporting smaller households on fewer resources, often partly supported by remittances from migrant partners. Without this kind of understanding, simple comparisons of how much male and female farmers produce can be misleading and unhelpful.

Most of the evaluations report on programme approaches to measuring and monitoring VfM – generally concluding these were unsatisfactory. LEAD is reported to be inconsistent on reporting against VfM, without clear systems in place to capture VfM data. FTESA is criticised for lacking targets for its economy and equity indicators, and over-emphasising economy and efficiency dimensions. The VfM strategy for MADE is reported to be not conducive to guiding programme operations, with high-level indicators that are difficult to use to prioritise and make strategic decisions. While ELAN's understanding of VfM principles is considered impressive, its framework is not as informative or effective as it could be – generating an excess of data without corresponding analysis. The evaluations also take issue with the VfM indicators that have been developed: the AECF evaluation questions leverage ratios where funds recorded as matching funds are sometimes already allocated for the project or were obtained through a process parallel to the AECF funding; the ELAN evaluation notes that the log frame target income increase of £30 is marginal, even for the very poor. There is no consensus across the evaluations as to what level of administrative cost is acceptable (and no consistency in calculating the cost); similarly with costs per beneficiary reached, the variations between programme approaches and circumstances are so great, and the level of service delivered to each beneficiary so different, that these high level indicators may not be very informative – and they are only as accurate as the data on beneficiary numbers.

The evaluations communicate a general dissatisfaction with programme approaches and available data on VfM, which make it difficult for the evaluations to give a confident assessment of whether the programme can be said to be providing value for money or not. The ELAN evaluation does note that subsequent to the evaluation the VfM framework has been improved – this could potentially provide some insights into meaningful ways to measure and report VfM for agriculture and M4P programmes.

j. Advantages and challenges of regional or multi-country approaches

The evaluations of the two regional programmes do not dwell on advantages and challenges of regional or multi-country approaches vis-à-vis single country approaches; and no lessons stand out from this learning review that are specific to the regional coverage of the FTESA and WAFM programmes. This highlights a potential area for future research into how regional or multi-country approaches compare with single country approaches.

Some methodological challenges that emerge from the regional approaches are:

- A critical issue is the difficulty of measuring cross-border trade and assessing the extent to which direct beneficiaries' (farmers') produce crosses the border. Traders and other buyers usually take the produce across borders, not farmers. For crops in tight value chains (e.g. tea, coffee, cocoa) tracing is relatively straightforward. But in loose value chains and where trade combines informal and organised routes, it is more difficult to trace the final destination.
- Both evaluations report significant difficulties with the regional model, and conclude that impact was limited. However analysis is limited and it is not clear that these difficulties were specific to the regional model. The common theory of change was that increased production and better marketing would lead to increased inter-regional trade in food grains and less fluctuation in market prices. Output side interventions in storage, grading and contract farming were also a key element in both projects. Farmer uptake was below what had been expected, indicating that the potential benefits had been over-estimated. More analysis (by the programme or the evaluation) might have explained how projections for benefits to farmers were calculated.
- Both FTESA and WAFM included a policy component, intended to help improve the enabling environment by addressing policy restrictions that hinder cross-border trade. Targeted activities appear to have been moderately successful in the case of FTESA, but mostly blocked or stalled at the final hurdle. (Follow-up research could highlight useful lessons if any of these have turned out to be successful in the longer run.) The WAFM evaluation finds that the Policy Facility was poorly linked to Challenge Fund activities, which undermined its effectiveness from the start. The FTESA evaluation advocates continuous and multi-dimensional policy engagement for more systemic effect; the WAFM evaluation recommends policy interventions that target specific policy challenges affecting the Challenge Fund projects (as was the case under FTESA).
- Both FTESA and WAFM used challenge fund mechanisms in an attempt to broaden their reach (as does the AECF), but neither evaluation analyses the pros and cons of the challenge fund modality vis-à-vis other modalities. The WAFM evaluation does note that 'there was an

intention that Challenge Fund grantees should demonstrate a cross-border element of their existing business model … this appears not to have been a priority in identifying grantees for implementation'. This indicates scope for further research to better understand the benefits and limitations of challenge fund modalities for agriculture and M4P programmes, through a more forensic investigation of where in the challenge fund process activities start to diverge from high-level strategic intentions. A starting point and good indicator of challenge fund effectiveness in project selection would be the volume and quality of applications submitted. This would shed light on whether the fund's potential market is large enough and has sufficient capacity to ensure a competitive process, enabling the fund manager to select from the best. The risk is that availability of applicants will determine what the programme does, as highlighted by the IMSAR review: 'Partner selection has effectively shaped the initial interventions, determining both what might be addressed as well as the extent of potential change'. For challenge funds, such as those run by FTESA and WAFM, this is unavoidably true, since the programme is shaped by the relevance and quality of the project proposals it receives.

2. Coherence of outcome and impact statements with the CFA

The CFA, published in 2015, provides a conceptual framework to guide DFID's approach to agriculture and the agrifood sector. Its case for intervention is broader than but not substantially different from the M4P logic:

> *broad-based agricultural transformations that leave no one behind require public interventions to address coordination and market failures in addition to investing in public goods and creating an enabling environment.*

It sets out a twin strategy to promote agricultural transformation focused on i) commercialisation and agroindustry development, to create jobs and raise incomes and ii) facilitating a long-term rural transition from subsistence agriculture to off-farm job opportunities as these emerge. This is presented in terms of three livelihood strategies of the rural poor:

1. **Hanging in:** Households that depend on subsistence agriculture for food and security and as a vital safety net. This calls for investment in subsistence agriculture to enable poor households to make the most of existing assets alongside other interventions to support their mobility and build human capacity.
2. **Stepping out** where poor people are able to access jobs through better roads, skills and improved health and transfers as a result of long-term investment in labour-intensive growth in manufacturing and services.
3. **Stepping up** where some in the farming community gain the scale needed for commercial operation through agricultural transformation which promotes agricultural commercialisation and agroindustry development to raise farm incomes, create jobs and lower relative food prices.

The CFA provides a more long-term, dynamic view that requires development programmes to identify interventions which meet the different needs of each farmer group. The contrast is with M4P's main focus on the market and on market actors.

DFID's agriculture programmes and the CFA

Publication of the CFA came after the design of all of the programmes covered in this review, and to an extent the findings from Section 1 highlight key insights from the CFA as summarised in the concluding key implications for agriculture interventions, in particular:

- Agricultural development requires a **differentiated approach** tailored to the opportunities and challenges of different categories of farmers, agro-climatic and geographical zones, and value chains;
- Agribusiness and value chain development require careful analysis to identify which models and markets hold the most promise for development. **The greatest opportunities for development impact may lie outside the value chains and geographical areas that tend to receive most attention**.
- DFID's agriculture programmes should **improve how they track impacts on women** and other marginalised groups, while supporting concrete action to create equal opportunities in agriculture for women and women and marginalised groups.

Comparing programme log frames with the CFA

Annex 1 brings together the log frame impacts, outcomes and outputs of the eight programmes, for comparison with the broad directions set out in the CFA. The overview highlights a high degree of consistency across the programmes, which is largely a function of the common use of the M4P approach:

- At impact level, six programmes target poverty reduction and livelihoods, while the two regional programmes target food security.
- Outcome-level statements focus on better functioning, more accessible markets – with the exception of LEAD, which focuses on smallholder incomes. Different drivers or corollary results are identified as second outcomes: drivers include cross-border value chains (FTESA) and innovation (IMSAR) while results include resilience (PropCom) and inclusion of victims of trafficking (MADE).

The log frames present some challenges in terms of their internal coherence, their underlying assumptions and (lack of) analysis, as reflected in the lessons identified in Section 1; and challenges in the understanding of the differences between impacts, outcomes and outputs. An overview of these challenges is presented in Annex 2.

It would be a mistake to try to distil from the CFA a standard model for commercial agriculture programming, to serve as a gold standard against which to evaluate agriculture programmes. In order to understand how well the programmes under review align with CFA priorities, however, we have unpacked some of its key conceptual underpinnings to guide assessment of impact and outcomes.

- At **impact** level the CFA has three interconnected goals:
  - Economic growth and poverty reduction
  - Food security and improved nutrition
  - Sustainable food systems
- At **outcome** level, the CFA conceptualises change in terms of:
  - Rural transitions – facilitating mobility out of the farm economy and from rural to urban areas while building the resilience of smallholders and raising returns to existing farm assets (from 'hanging in' to 'stepping out');
  - Agricultural transformation – agribusiness investment and value chain development downstream from the farm to enable diversification out of primary commodities and value addition ('stepping up');
  - Getting the basics right – investment in public goods and an improved investment climate to build a supportive enabling environment for agricultural transformation and rural transition.

This summary illustrates that the set of eight agriculture programmes fits comfortably within the CFA. Six programmes fit within the economic growth and poverty reduction goal while the other two address food security and improved nutrition. All programmes are clustered within the agricultural transformation and 'stepping up' strategy, underpinned by the focus of M4P and market systems interventions on value chain development and greater participation by poorer farmers.

At the level of log frame impact and outcome statements, therefore, the set of programmes is consistent with the CFA. However, this is largely by virtue of phrasing and the general level at which these statements are made (see Annex 2) – it does not necessarily follow that the programmes are more meaningfully coherent with the CFA at lower levels of the logic model, in terms of their theory of change assumptions and the implications for how programmes are designed. The CFA highlights five key questions to guide the programming of agricultural transformation interventions, some of which resonate with evaluation findings and lessons identified in Section 1:

- Which scale of farms and scale of agribusinesses should be prioritised? – with a focus on linking small-scale commercial farms to agribusiness and their value chains and supporting such linkages with existing commercial and emergent commercial small-scale farmers. What scale of agribusiness has the greatest potential contribution to make to inclusive transformation processes?
- What institutional arrangements should be promoted to ensure agribusiness investments benefit the largest number of small-scale farms? i) vertical coordination arrangements between actors at different stages of the value chain or ii) horizontal coordination between players at the same stage of the value chain?
- Which value chains or related agribusiness investments are likely to benefit the largest number of poor people in the long run, as producers, workers or consumers?
- What are the main risks from prospective agribusiness investments that need to be prevented or mitigated? Land transactions stand out as the single biggest, material risk.
- What is the case for intervening? – to what extent is private sector investment leveraged by public intervention additional?

Based on the evidence from the evaluations reviewed, we would add an emphasis on a strong understanding of farmers' resource endowment and strategies. For example:

Hanging in. One modern input can be found in wide use in some of the poorest parts of the world. From rural Nicaragua to central Africa, herbicide is widely used in what can be best described as a smallholder form of minimal tillage. In effect, herbicide applied with knapsack sprayers has been substituted for slash and burn. This is a dominant pattern in communities where much of the male work force has stepped out to work in the cities or overseas, whilst women provide for family nutritional needs (commonly growing maize, millet, cassava, and beans as appropriate to their environment) with the smallest possible labour input.

Stepping up. M4P programmes are aimed at helping smallholders to increase productivity and add value through storage and grading to increase marketed output. In this respect, they are coherent with one part of the CFA strategy: stepping up.

With some exceptions, programmes have focussed on promoting the use of inputs which raise yield – land saving technologies. As the CFA recognises, agrarian development has rarely taken place without some degree of consolidation into larger farm units which can benefit from labour saving inputs, mechanisation above all. Land is still relatively abundant in many regions of Africa and it is perfectly possible that some farmers will step up by farming larger areas with little or no change in yield; especially in dryland areas. It is relevant that some programmes in the current portfolio include interventions for partners to clear new land for farmers.

Intensification or extension is just one of the strategic choices stepping up farmers will have to make. They will also need to select the optimum combination of crop and livestock enterprises to make use of their resources; and they will need to decide which are their best markets. The programmes reviewed have encouraged farmers to sell to larger commercial enterprises which mainly serve modern sector markets: schools, hospitals, supermarkets, etc. In most African nations this sector is still quite small. Informal urban markets for ungraded produce are much larger and they are often growing faster as well. This may well have been a factor behind FTESA and WAFM's lack of success with grading and storage interventions.

These brief points are to emphasise that if CFA is to succeed, programmes will need to put significantly greater effort and resources into research in order to understand the needs of specific farmer groups in specific contexts and design interventions which meet those needs.

In processes of agrarian change, equitable consolidation of landholdings is a critical social and political issue. The essential is that farmers who give up their land get fair value for it. They may do this by selling it, but also by renting or sharecropping it out. The income they get from this can be a key part of a hanging in strategy. None of the programmes in the portfolio has included interventions in land markets. DFID and others have supported major land titling programmes in many African countries. However, titling is only a first step towards the development of a fair market in land.[5] In most cases informal land markets anticipate the development of formal processes. In practice, interventions to cover the gap between formal titling and land transactions are likely to be an important part of the CFA strategy.

The CPAR notes access to finance as a significant gap in DFID programming for agriculture. Some programmes in the portfolio reviewed here included financial interventions. Those linked to storage have not been successful. Other lending models have shown the common tendency for credit to be diverted to other uses and repaid from non-agricultural sources. Transaction costs in agricultural finance are high as are the risks of moral hazard. The clearest lesson from the current portfolio is to confirm the importance of putting adequate effort into researching and designing robust interventions that meet farmers' real needs.

3. Advantages and disadvantages of evaluation methods used

The set of evaluations reviewed includes a mix of mid-term and final evaluations, some of these conducted as part of a longitudinal evaluation process; plus one interim review and the two IMSAR topic-focused reviews. These adopt formative or summative approaches depending on the stage in the programme, and apply a generally standard set of evaluation methodologies. Table 5 provides a summary, based on the intentions stated in the methodology section of the evaluation reports.

[5] Some DFID titling programmes do include components to encourage the development of land sale and tenancy arrangements. LIFT Ethiopia is an example.

Table 5: Summary of evaluation types and approaches

	Stage	Type and focus*	Use of survey/ other means of attribution	Approach/ methods	OECD DAC & other criteria
FTESA	Mid-term	Formative	Online grantee survey	Theory-based Thematic studies Realist baseline case evaluation synthesis	Relevance Effectiveness Efficiency Synergies Sustainability Replicability Cross-cutting Governance
FTESA	Final	Summative Performance and impact	Baseline and endline quantitative household surveys for two grants including control group Contribution analysis	Theory-based Case studies Thematic studies Realist enquiry	Relevance Effectiveness Impact Efficiency Replicability Synergies Cross-cutting Sustainability
WAFM	Mid-term	*Formative* Process	Contribution analysis	Theory based Realist baseline case study	Relevance Effectiveness Efficiency Impact
Prop Com	Final	*Summative* Impact	Baseline and endline survey in two primary markets including control group	Theory based	Relevance Sustainability Effectiveness Impact Efficiency
MADE review	Interim	*Summative* Performance and impact	MADE surveys of changes in behaviour, productivity and income Benchmarking		
ELAN MTE	Mid-term	*Formative*			Relevance Effectiveness Efficiency
LEAD	Mid-term	*Summative*	BRAC household survey at baseline and endline including counterfactual	Theory-based	Effectiveness Efficiency Relevance Sustainability Impact
IMSAR Q1 **IMSAR Q2**	Y1 review	*Formative* Process review	n/a		Relevance Efficiency Effectiveness Impact Sustainability
AECF	MTE	*Formative* Process and impact	Applicant survey Benchmarking		Relevance Efficiency Effectiveness Sustainability

* This has been inferred where in italics, but not stated in the report

Generation of primary data

The evaluations rely to a varying extent on monitoring data generated by the implementing partner. Results data is generally the responsibility of the programme implementer but in the case of some longitudinal evaluations the evaluator has been responsible for a degree of verification of the reported results. For four evaluations that benefit from a quantitative baseline-endline survey, two were implemented by the evaluator and two by the implementing partner. For most evaluations, the evaluator expected the implementing partner to report against VfM metrics, although some supplementary construction of VfM metrics was done by the evaluator in two cases. Table 6 summarises the division of responsibilities for data gathering.

Table 6: Division of responsibilities for data gathering

	Results monitoring data	Survey	VfM data
FTESA FTESA	Grantees PMU verification	Evaluator	PMU (limited data) Evaluator benchmarking
WAFM	Implementing partner MRM verification by evaluator		Evaluator
Prop Com	Implementing partner	Evaluator	Implementing partner? (VfM annex not available)
MADE review	Implementing partner	Implementing partner	Implementing partner
ELAN MTE	Implementing partner MRM verification by evaluator		Implementing partner
LEAD	Implementing partner	Implementing partner	Implementing partner supplemented by evaluator
IMSAR Q1 IMSAR Q2	n/a	n/a	n/a
AECF	Fund Manager & grantees MRM verification by evaluator	Evaluator	Fund Manager Ratios prepared by evaluator and FM

Evaluations cannot depend on availability or quality of programme data. Dependence by the evaluation on partner reporting – and lack of confidence in reported results – has been an overriding theme. Data (commonly VfM data) has been absent, has been incomplete or of low quality. The MADE evaluation comments, ‘The programme’s quantitative results require validation. M&E systems are weak and evaluation reporting needs to be substantially strengthened’.

Where there has been verification by the implementing partner, this has not necessarily provided assurance for evaluators. The AECF evaluation notes ‘MRM reports seemed to be just accepting the grantee information and statistics – rather than taking an objective view’; and the ELAN evaluation: ‘As part of the MTE process, a more in-depth review of reported results for specific interventions, including through primary data collection, raised concerns with measurement that had not been revealed during the verification process.’ As a desk exercise, the verification was, in effect, a test that the data was complete, internally consistent and within expected bounds. This did not necessarily mean it was accurate. In their fieldwork the ELAN evaluators found their primary data collection cast doubt on the reliability of data that was assessed as adequate by the desk-based verification exercise. Their inability to be confident in programme data led them to suggest abandoning any further verification, to be replaced by greater emphasis on primary data collection.

The ELAN response is one strategy to mitigate poor data quality. Other evaluations base their assessments quite largely on interview and focus group evidence, arguing that triangulation across these sources would sufficiently underpin quality. Even this has its limitations. As the FTESA final evaluation puts it, 'The ability to triangulate and synthesise effectively across a range of data tools and levels was constrained by the presence of some gaps in information. In some cases, the team relies on presenting examples within the narrative to substantiate findings'. In a number of cases, evaluators resort to including quotes from single sources.

The WAFM evaluation highlights **underlying causes of resource availability and grantee capacity vis-à-vis the reporting burden**: 'The programme relies significantly on its grantees to collect and report programme data (partially because of resourcing issues) … However, it is a task that not all grantees are necessarily sufficiently equipped for'. Programmes which work through partners, be they commercial organisations, NGOs or government ministries, must largely depend on those partners to report programme achievements; and the incentives for different types of partner to comply with reporting requirements vary. Setting the data requirement at a realistic level, i.e. one which the partner can and is willing to fulfil, is a common challenge. The WAFM evaluation describes how programme managers, perhaps in response to DFID requirements, pushed to expand the list of indicators while partners and grantees felt that 'programme requests for data were too frequent and too complicated, and took up too much company time'. As a result reporting templates were not always completed.

This highlights gaps in expectations in terms of i) what can be achieved with partner reported data (which can be expected to be of variable quality) and ii) the extent of primary data collection that can be done by the evaluation with available resources. Ideally partner reported data and primary data collection by the evaluator would overlap, so as to enable triangulation and underpin confidence in the evidence and findings. In practice it seems that there continue to be evidence gaps, indicating the evaluation data collection strategies are not well adapted to the coverage and quality of results monitoring that exists.

Beneficiary numbers and income impact

A particular challenge has been to convincingly report against high-level indicators for beneficiary numbers and income impact. Any robust estimate of programme impact must start from a statement of the number of beneficiaries reached and the estimated income increase which resulted, however the evaluations demonstrate general weakness in being able to present independently verified data on beneficiary numbers and credible estimates of net attributable income change (NAIC).

A particular gap that complicates the understanding of impact is that the evaluations did not, in general, present a clear description of the pre-intervention situation – the programme start line – which might have taken the form of a set of simple indicators such as:

- For beneficiaries: number of farmers, current production, current input use
- For commercial partners: current levels of sales in the programme area, existing networks of dealers
- For markets: number of actors in different roles, estimates of their level of business

Three evaluations present a full statement of beneficiary numbers and income impact: PropCom, MADE and AECF. Both the PropCom and MADE reports caveat their findings with the statement that the data has been provided by the programme and not been independently verified; and raise doubts about how results have been calculated (or lack of visibility on the methods used). The AECF evaluation also includes the proviso that 'the total number of beneficiaries and net benefits will in most cases always be an estimate which is unavoidably surrounded by a margin-of-error'. The fact that the evaluation nevertheless uses the data to demonstrate impact indicates a judgement that this estimate is 'good enough' for the purposes of the evaluation: we pick up on this theme under Question 5.

The other evaluation reports provide few firm figures on beneficiary numbers and income increases:

- **LEAD:** no beneficiary numbers and only scattered statements on the income increase per beneficiary;
- **WAFM:** beneficiaries per grantee, with a total of 71,000 (68,000 from two grantees) and 'reports' of increases in smallholder incomes from three of the six grantee case studies;
- **FTESA:** VfM analysis includes 168,913 small farmers engaged (annexed log frame shows no actual figures after Year 1) and reports 'very few recorded examples of increasing farmer incomes';
- **ELAN**[6] reports some 12,000 'indirect' beneficiaries, and reported outreach of 400,660 'direct' beneficiaries with increased income (annexed log frame shows no actual figures after Year 1), but also that 'it seems unlikely that MSCs across ÉLAN's sectors will be sufficient to meet the NAIC impact targets set out in the log frame.

Methodological approach

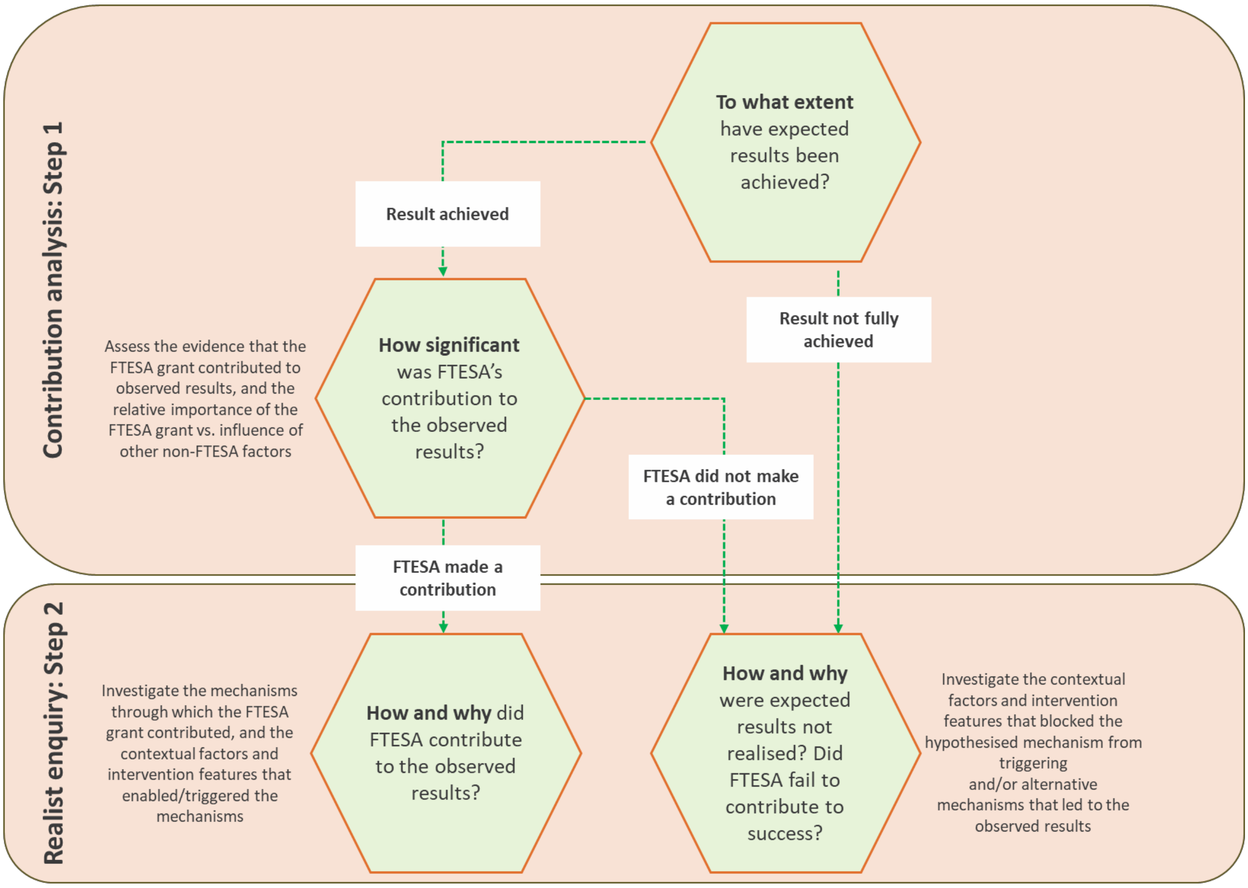
Methodological approaches are broadly standard across the set of evaluations. Most apply theory-based methods, taking the theory of change as the starting point for identifying or expanding on a set of evaluation questions, with data sources and data collection methods identified for each within an evaluation matrix. They take a mixed methods approach, gathering qualitative and quantitative data through document review, interviews and focus group discussions – with some additionally including some form of survey – and almost all report against OECD DAC evaluation criteria.

Three evaluations – the two FTESA evaluations and the WAFM MTE (all carried out by the same company) – describe a **realist enquiry approach**, which has been used in combination with **contribution analysis**. Realist approaches highlight the influence of context on programme outcomes (posing the 'what works where?' question). They begin by clarifying the 'programme theory' (how the programme expects to effect change), the mechanisms that are likely to operate, the contexts in which they might operate, and the outcomes that will be observed if they operate as expected (the CMO configuration). Detailed methodologies are provided in the FTESA Case Evaluation Synthesis report; and the WAFM Baseline Case Study Synthesis report.

The advantage of applying a realist approach to organise and make sense of evaluative material is highlighted by the EQUALS quality assessment of the FTESA final evaluation, which notes that its section on findings is 'a very thorough and clear synthesis of the evidence emerging from the modules, structured consistently in line with the realist evaluation approach, with summary findings to introduce each EQ. The aggregation of a large volume of evidence is impressive.' The Case Evaluation Synthesis report provides an overarching illustration of the methodology (Figure 3) as well as an evaluation matrix that sets out, against each evaluation question, intervention-context-mechanism-outcomes (ICMOs) and realist enquiry questions, as well as the programme theories (Annex 8.3 to the report).

[6] ELAN was a mid-term evaluation and arguably too early to expect full impact data.

Figure 3: FTESA methodology combining contribution analysis and realist enquiry



The EQUALS review of the WAFM evaluation finds that the evaluation was less successful in bringing together findings and conclusions, with limited visibility in terms of being able to trace evidence through the analysis and into findings, conclusions and recommendations. This is not explicitly attributed to how realist enquiry was applied. However, the reviewer highlights that although the evaluation draws on realist thinking in terms of acknowledging the importance of context, the mechanisms set out in the identified CMO configurations are not strictly causal forces or powers as traditionally understood in realist thinking.

The purpose of the **baseline case evaluations** for FTESA and WAFM is to explore the underlying theories for a purposeful sample of grants for further testing in later stages of the evaluation, and explore early findings about how and why the programme works (and does not work). The final round of qualitative case studies is intended to test the baseline CMO theories to determine whether or not these theories explain how change has occurred across the different grants.

The WAFM baseline case evaluation synthesis report describes the methodology for arriving at the final set of CMOs and presents findings and cross-case comparisons for six case studies of WAFM grants to agribusinesses that set out the constraints grantees and smallholder farmers faced prior to the WAFM intervention, how the WAFM-funded grantee models have attempted to address these constraints, and what challenges remain. The findings are structured according to the key stages in the staple food production and marketing cycle, with references to CMOs interspersed throughout.

The evaluation reports and annexes for both FTESA and WAFM illustrate the usefulness of the methodology for systematically exploring causality and the how and why of change in a given context – that is, for addressing theory-based questions in the evaluation matrix. The WAFM MTE notes that evidence from the process and realist evaluations is then synthesised through thematic analysis, using the OECD DAC evaluation criteria.

The advantages associated with the realist methodology of cataloguing and categorising evidence to enable the systematic identification of findings are offset by disadvantages which the WAFM baseline case evaluation synthesis report highlights in its summary of limitations and challenges:

- The process of doing two types of synthesis in parallel, combining realist evaluation with a programme evaluation, was extremely time consuming. Analysis generated an unwieldy number of CMOs that had to be sifted and rationalised.
- There were further limitations relating to the abstract nature of both realist evaluation and ToC thinking. There was a lack of enthusiasm among participants (particularly private sector actors) to reflect on how the ToC worked.
- The analysis found limited evidence emerging from interviews at higher levels of abstraction from the activities themselves (e.g. income level), where links between the Cs, Ms and Os are typically distant and fuzzy. Hence, at this level, many of the CMOs remain hypothetical.
- There is a risk of positive bias in data generated through programme and grantee staff, especially in relation to linking implementation to outcomes through mechanisms at the institutional level.

In summary, the use of realist enquiry has strengthened confidence in the evidence base and facilitated the systematic organisation of a large volume of data against the evaluation questions. However, this advantage has come at a cost in terms of resource intensity and greater complexity in the analysis in synthesising findings from the realist evaluation with findings from process or performance evaluation.

Addressing attribution and contribution

Four evaluations use quantitative survey data to help establish the role of the intervention in bringing about the observed change.

- **FTESA:** The evaluation team conducted baseline and endline household surveys for two (of 20) grants (two of six case studies) to assess impact on beneficiaries, including control groups. For one of these, the endline survey was downscaled to a monitoring survey (interviewing only beneficiary farmers in the intervention group and not the baseline control group) as a planned warehouse was not operational in time. Survey data was analysed using before-and-after and difference-in-difference methods.
- **PropCom:** The evaluation conducted quantitative baseline and endline surveys to capture the changes and intervention impacts in two of its primary markets, with control groups.
- **LEAD:** The evaluation used data from a 2016 household survey conducted by the implementing partner BRAC, as a follow-up to a baseline survey, using control and treatment selections within the same sampled area, and a 2015 adoption survey. The evaluator reviewed the quality of the survey process, report and data and concluded that the general quality of the data was good with some minor problems. Limitations included the fact that the data was not disaggregated according to gender or location. The report notes that the evaluation team was not given access to the complete dataset because of time constraints.
- **MADE:** The evaluation used data gathered by MADE through surveys that measure changes in behaviour, productivity, and income. Survey data is gathered in 'results reference sheets' that are analysed to estimate the number of farmers changing behaviour, increasing productivity and increasing income. This is clearly less rigorous than the above surveys and the evaluation articulates concerns about inbuilt assumptions and the quality of the process and results. Notably, it suffers from the absence of a control group to underpin attribution.

The availability of survey data – where there is confidence in the data – helps to establish confidence in evaluation findings; and this is reflected in the PropCom evaluation where findings are presented with an assessment of strength of evidence which, in the case of impact, is rated 'conclusive'. None of the evaluations, however, gives great prominence to survey results. This may reflect unarticulated concerns about the reliability of the data or its interpretation. It may also reflect the fact that findings from surveys conducted on two out of 20 grants, or in two out of a larger number of primary markets (it is not clear how many), cannot easily be extrapolated to other grants and other markets. If this is the case, this would raise value for money questions about the survey design choices.

What is also surprising is the general absence of counterfactual analysis and assessment of attribution, given the investment in including control groups. Results are generally reported in terms of the changes experienced by target beneficiaries, without comparison with the 'without treatment' scenario. The inference is that the data does not support attribution: for instance, the LEAD report presents the following conclusion on impact: 'Provision of information regarding good husbandry and agricultural practices has resulted in increased yields allowing for increased consumption of maize within the household and increased prices of sold poultry, with both outcomes adding to increased household resilience.' In fact, review of the survey results annexed shows that treatment farmers used more fertiliser than the control but there was no statistically significant difference in any output measure: production, yield, income or sales. Although not as clear, the poultry results do not seem to be different. Surveys are, therefore, finding no evidence of attributable impact either because there is no causality (an evaluation finding which should be reported more prominently) or because the survey design is underpowered and the data cannot support conclusive findings.

In the case of FTESA, where results have been somewhat meagre, there is less of a case for needing to establish attribution.

The one evaluation that does refer to the control group uses this evidence inconsistently: the ELAN evaluation states, 'All the interventions on which data are available indicate that incomes have increased'. However, it then reports that only one of the two surveys showed a significant increase in treatment farm incomes relative to the control group. 'In the latter intervention … it was found that on average net incomes of Babban Gona farmers nearly doubled, but comparison farmers experienced a similar increase'. It would seem that other, less rigorous data was taken to outweigh this survey result in concluding that the intervention had made a positive impact. The evaluator goes on to quote without qualification the doubling of maize farmer income as evidence of project impact.

Closer analysis would be needed (the PropCom survey results are not available) to understand whether the use of control groups have helped to build evidence for attribution – and to address broader questions about the suitability of M4P programmes to proper estimation of counterfactuals. But if survey results are so inconclusive in a sector where it is relatively easy to count things, **surveys and impact evaluations need to be more carefully designed to ensure they are able to answer evaluation questions with a sufficient level of confidence to ensure VfM.**

Two other evaluations discuss the role of a counterfactual, although they do not go into detail about what form this might take. Both the AECF and ELAN evaluations criticise implementing partners for not establishing counterfactual data gathering

4. Consistency of evidence for common evaluation questions

We collated the evaluation questions from across the seven programmes evaluated to draw out the most common areas of inquiry (paraphrased in Table 7). Below we assess the extent to which the evaluations converged or diverged in terms of their response.

Table 7: Most common evaluation questions

Common evaluation question	FTESA	Prop Com	MADE	WAFM	ELAN	LEAD	AECF
Is the programme delivering VfM?	✓	✓		✓	✓	✓	✓
To what extent is the programme contributing to or likely to contribute to systemic change?	✓	✓	✓	✓			
How effective has the programme been in delivering log frame outcomes and impacts?		✓	✓		✓		✓
To what extent does programme governance and management underpin delivery?				✓	✓	✓	✓
How well is gender integrated into the programme?	✓	✓				✓	
To what extent has improved availability and use of inputs helped deliver expected results?	✓			✓		✓	
Does the programme have robust results monitoring systems in place?					✓	✓	✓
Are results relevant to beneficiary needs, including marginalised groups?		✓		✓		✓	

PropCom dropped the VfM question in its final evaluation report, but it is included here given that VfM was the subject of a dedicated analysis. MADE is included on the basis of its stated objectives – it did not articulate formal evaluation questions

Is the programme delivering VfM?

As shown in Table 6, the evaluations mostly rely on the implementing partner to generate VfM data and identify indicators; although some did some supplemental work to construct VfM ratios and to benchmark to other comparable programmes. (The FTESA evaluators worked with the PMU to develop its VfM metrics midway through the programme.) There is a general dissatisfaction with the quality and coverage of VfM reported data and analysis, and evaluations are consequently reluctant to make definitive conclusions on VfM, or conclusions that are not heavily qualified.

The PropCom evaluation, as noted, had VfM as one of its formal evaluation questions, but did not include the analysis or conclusion in its final report

Annex 3 summarises the VfM metrics referenced across five evaluations, according to the 3Es framework (plus equity for FTESA and WAFM). The table includes additional metrics calculated by the evaluation, as well as metrics identified but not reported against by the implementing partner. It also presents the metrics as presented by the evaluation or implementing partner. Three evaluations benchmark the indicators against similar programmes: the FTESA evaluation takes WAFM and AECF as benchmarks, the WAFM evaluation takes FTESA, and the AECF evaluation benchmarks against the Enterprise Challenge Fund (ECF) and a private equity fund, AfricInvest. The LEAD evaluation makes generic comparisons, without identifying comparators. Only the FTESA evaluation considers how VfM changes over time.

In general, several of the common indicators rely on available and relatively verifiable data – but what they actually measure varies so considerably that any comparison or benchmarking becomes difficult.

This leaves open the question of deciding on a target value at which VfM is considered to have been delivered.

Economy

Economy indicators focus on the management and administration costs of delivering the overall programme, including personnel and TA costs. These are calculated differently across the different evaluations: FTESA calculates a broad fund management ratio, as well as an administrative cost ratio while ELAN and LEAD focus on personnel costs, distinguishing between the costs of national and international experts.

Figures range from an overhead ratio of 15.6% for LEAD to administrative cost ratios as high as 32% (FTESA) and 27% for AECF. The FTESA evaluator considered 32% to be high compared to other programmes, where the AECF evaluator felt 27% was acceptable. This serves to underline the danger of using ratios in this way. The line between delivery and overhead costs is a grey one and programmes like FTESA, which have policy facilities and disparate partners, are likely to encounter higher costs than a straight investment fund like AECF. That said, 32% is a large share of the total fund.

Efficiency

Most evaluations look at some variation of the cost per beneficiary. These figures range from £17 for WAFM to £78 for LEAD. As with the administrative cost ratios, the variations between programme approaches and circumstances are so great, and the level of service delivered to each beneficiary so different, that these high level indicators may not be very informative. (And they are only as accurate as the data on beneficiary numbers.)

Another common metric is leverage. A key goal of M4P is to stimulate matching investment by programme partners, either from their own resources or from funds drawn in to match the DFID contribution. Reported leverage ratios range from 1:1.44 for WAFM to 1:5.95 for AECF. The AECF evaluation highlights 'the fact that funds recorded as matching funds are sometimes already allocated for the project or were obtained through a process parallel to the AECF funding and therefore the catalytic effect cannot be proven'.

This raises the question of additionality, which is considered in the WAFM, ELAN and AECF evaluations, and one of the IMSAR reviews. The WAFM evaluation recommends a detailed additionality screening which could have helped to incentivise innovation and drive impact. Assessing additionality is a central, if difficult, task that needs to take account of a range of intersecting factors. As AECF points out 'additionality presents a trade-off with other criteria, for example a project which is highly additional given the fact that the grantee does not avail of the funds, may for the same reason have a large chance to be non-sustainable and therefore have no impact. A very large leverage through substantial matching funds may at the same time indicate a low additionality'.

Two evaluations use composite indicators for efficiency: LEAD's cost efficiency measure factors in costs per unit for the number of partners continuing 12 months after pilot end, partner spend, number of poor people changing business practice, number of competitors copying or expanding innovations and number of non-competitors changing business practice. AECF assigns an 'efficiency score', but the methodology for this is not provided.

Effectiveness

There is no common approach to assessing effectiveness across the evaluations. ProCom, AECF, ELAN and LEAD use different measures of cost-effectiveness – and only the first two calculate this in terms of the benefit per beneficiary divided by the cost per beneficiary which is given as 1:1.19 for ProCom and 1:1.06 for AECF. The estimates of benefits appear to be based on a single year's additional income and, as the PropCom evaluation states, it assumes the total gain is attributable to the programme. The rigour of these estimations is therefore questionable.

The evaluations highlight an important point. A cost-benefit ratio close to 1:1 implies that the programme could have had the same impact on beneficiary incomes by distributing the money direct to smallholders, and an even greater impact if administrative costs are reduced. The ELAN evaluation makes a related point – that the log frame target income increase of £30 is marginal, even for the very poor, amounting to less than two cents a day per capita for a five-person household.

Equity

Only the FTESA and WAFM programmes included VfM indicators for equity: both measured the percentage of women among smallholders engaged, while the FTESA evaluation also calculated the cost of female outreach.

To what extent is the programme contributing to or likely to contribute to systemic change?

Of the four evaluations that ask about systemic change, one is an MTE and one an interim review. All comment that it is either too early, or the duration of the programme is too short, to be able to confidently assess the likelihood for systemic change. Where systemic change is noted, this is in the two legacy markets of the PropCom programme – and this is qualified by 'given sufficient time'. There is common use across three evaluations of the AAER framework for assessing systemic change – looking for examples of the 'expand' and 'respond' stages (of which there has been limited evidence across the set of evaluations) – albeit with varying degrees of rigour: the MADE review is relatively superficial compared with the very systematic AAER approach by the FTESA evaluation. WAFM does not explicitly ask about systemic change but rather the improved functioning of staple food markets and catalysing policy change.

How effective has the programme been in delivering log frame outcomes and impacts?

Three evaluations ask an overarching question about the extent to which the programme was successful in delivering log frame outcomes and impacts; and the MADE evaluation has this as one of its objectives. While the AECF evaluation poses the question, it does not present its findings against the questions, so the top-line assessment of performance against the log frame is buried (or the analysis was not structured this way – it is not clear from the description of the methodology).

The PropCom and MADE evaluations are generally positive on log frame performance, the ELAN evaluation less so:

- **PropCom:** In seven of the nine evaluated interventions, incomes were shown to increase. In all of these seven interventions increased incomes also led to higher productivity. The increased provision of services across many of the interventions … demonstrate an increase in the functioning of rural market systems, with private sector partners committing their own funds to increase their reach and scope in a number of the interventions.
- **MADE:** MADE I has exceeded all its log frame targets (including outputs, outcomes and impact) in a very difficult operating environment.
- **ELAN:** Under the most plausible assumptions associated with existing interventions, it seems unlikely that MSCs across ÉLAN's sectors will be sufficient to meet the NAIC impact targets set out in the log frame.

To what extent does programme governance and management underpin delivery?

This is an important question for VfM, and a concern across all evaluations with the WAFM, ELAN, LEAD and AECF evaluations including it as a formal evaluation question. All four are considered to have 'room for improvement', for both internal and external reasons. ELAN, however, stands out as having driven positive results through flexible and responsive management: 'ÉLAN's processes … are efficient and effective. The initiation of some 170 interventions during the four years of implementation is testament to this highly productive and adaptive programme.' Across the four evaluations, the following issues are identified:

- **WAFM:** The programme has suffered from significant personnel 'churn', especially in the early, foundational phases, and this affected its early traction and resultant decision making and implementation.
- **ELAN:** Some flaws are evident in the design of certain partnerships, and ÉLAN's management processes have not always been effective in recognising and responding to these weaknesses and their consequences in a timely fashion.
- **LEAD:** The biggest organisational challenge LEAD faces is high staff turnover. There appears to be limited opportunity for local staff to initiate suggestions for change or improvement, entrenching service delivery rather than promoting market innovation.
- **AECF:** Challenges are identified in terms of lack of transparency on the ownership structure, unclear responsibilities, and lengthy grantee selection processes.

How well is gender integrated into the programme?

All programmes were reporting gender disaggregated data to some extent, and were able to show female participation, albeit below target in some cases. However, there is limited accompanying analysis:

- **FTESA (MTE):** Gender is largely absent from grantee strategies and, when mentioned, it focuses on women participating via membership to women-only collectives. In grantee results, many of the grantees report on gender-disaggregated data, however women appear to be under-represented with men benefiting disproportionately. (The gender question was dropped for the final evaluation of FTESA.)
- **PropCom:** The programme significantly increased its reach to women and girls … On the other hand, the programme fell short of its log frame target of reaching 250,000 women … it is worth noting that there appears to be no rational explanation in the business case for the significant increase in this target, possibly failing to take into account the economic role of women in northern Nigeria.
- **LEAD:** LEAD conducted gender analysis revealing the different experiences of men and women farmers, but it is not clear how this analysis is feeding into programming. Both men and women are involved in maize farming, while more women (and some older men) appear to be involved in poultry farming. It is not clear why women maize farmers earn less than their male counterparts do.

As noted in Section 1 there is a generally superficial treatment of gender in the evaluations as well as in the programmes.

To what extent has improved availability and use of inputs helped deliver expected results?

This is a specific question for the FTESA, WAFM and LEAD evaluations, with positive – but qualified – results reported for each of them:

- **FTESA:** Several cases present good evidence of improvements in the availability and use of inputs and farmers simultaneously applying good agricultural practices due to FTESA-funded activities, although the numbers reached are low, with supplies of inputs and numbers trained lower than expected … Where farmers have applied good agricultural practices and used improved inputs, productivity and quality has improve … There is limited evidence across the grants that these interventions have led to higher-level results (prices, sales, incomes), which is partly due to the stage of implementation and lack of success in finding markets.
- **WAFM:** While it is too early to draw firm conclusions against the EQ, the majority of grantees are providing inputs and services on credit and are increasing the number of farmers they are working with.

- **LEAD:** Maize groups report an increase in the application of GAP resulting in higher yields. Poultry groups report an adoption of use of tools for poultry feeding and watering, as well as the uptake of improved feeding and medical care practices resulting in higher prices for their birds. Provision of information regarding good husbandry and agricultural practices has resulted in increased yields allowing for increased consumption of maize within the household and increased prices of sold poultry, with both outcomes adding to increased household resilience.

Does the programme have robust results monitoring systems in place?

This is a question for ELAN, LEAD and AECF, with a number of reservations articulated – also reflecting the dependence of the evaluation on programme-generated monitoring data.

- **ELAN:** While there were some weaknesses in measurement, all indicators were rated with at least 'medium' confidence that reported results reflect reality. More in-depth review of reported results for specific interventions … raised concerns with measurement that had not been revealed during the verification process.
- **LEAD:** Reporting is regular and comprehensive with respect to reporting against the log frame. There is little reporting linking activities with outcomes or follow-up actions.
- **AECF:** While the majority of grantees provide report of sufficient quality, some lack either the competence or the will to provide good data. Monitoring information about beneficiaries is often new for grantees but appreciated in some instances. In many cases, self-monitoring carries the risk of too optimistic data reporting.

Are results relevant to beneficiary needs, including marginalised groups?

This is an evaluation question for PropCom, WAFM and LEAD, with findings reflecting the rather inconclusive findings noted in Section 1 on the trade-offs between results and reach.

- **PropCom:** if supply and demand side constraints are properly solved, this would allow private and public sector partners to provide products and services to address beneficiaries' needs.
- **WAFM:** The project does not collect data appropriate to judge the differential impact on youth or to divide beneficiaries by economic status. Such data is more difficult and costly to collect than for gender disaggregation.
- **LEAD:** LEAD's survey results indicate that participant farmers can be classified as subsistence smallholder farmers; however, it is not clear if or how LEAD is including the most marginalised within this demographic.

5. Opportunities for synthesising learning in the future

The CFA and consistency of high-level goals across the programmes evaluated certainly provide potential for synthesising across evaluations as demonstrated by the number of common evaluation questions (Section 4), the commonality of certain key issues (Section 1) and the applicability of standard, theory-based evaluation methods (Section 3). In the case of the set of evaluations reviewed, however, this scope is limited by:

- the inability to be confident in programme-generated monitoring data, and incomplete data collection (particularly across time in the case of income impact);
- insufficient comparability across 'comparators' and the inability to define a common threshold for 'good' performance against common metrics;
- the large and often unaddressed role played by context and externalities which qualifies the assessment of performance in each individual case.

This is compounded by conceptual uncertainty around how to approach questions relating to gender and, to a lesser extent, value for money – both by the programme and the evaluation. In some cases this has led to thin analysis and unsatisfactory conclusions (to the effect that there is not enough evidence, or the evidence is not good enough, to give an assessment).

Based on the findings emerging from the review, we have identified three avenues that hold potential for the purposes of synthesis, based on i) further developing and testing approaches applied by one of the evaluations reviewed; ii) identifying practical common metrics and iii) providing more explicit guidance on frameworks for assessing gender-related issues and VfM.

Approaches for further exploration

The FTESA evaluation's use of the **AAER framework** might be further explored for its potential for standardisation across other evaluations. Unfortunately, this was not provided for this review: we therefore recommend further analysis of how this was applied to establish whether it might provide a robust model that can be used in other evaluations.

The FTESA evaluation included two thematic studies, on EAGC and Farm Africa, which are reported to have applied the systems-level AAER framework in order to answer questions on systemic change and sustainability. Review of the Farm Africa study suggests that, in fact, only the 'adoption' dimension has been explored. It would be useful to be able to review the analytical framework applied (to the extent that it has been tailored from the Springfield model) and consider its applicability across a longer duration where the 'expand' and 'respond' results might be expected.

Use of common metrics

In practice, each individual evaluation arrives at a set of conclusions based on a unique configuration of factors, combined and weighted according to the expert judgement of the evaluation team. Evaluative choices are made that cannot be systematised for all evaluations, and depend on intelligent use of what evidence and tools are available. It is therefore not very practical to try to aggregate micro-level results (such as NAIC) where small differences in calculation methods can be amplified into large margins of error at the aggregate level. Further, the review has noted the common challenge of imposing data collection and reporting burdens on implementing partners with limited resources, capacity or will to do this well enough to be reliable.

An alternative would be to start with an assessment of how much data, at what level of granularity, is 'good enough' to be able to confidently answer high-level questions about whether a programme is making a positive difference, and whether this is enough vis-à-vis the level of investment (i.e. value for money). This calls for simple models that are robust to data variability within a relatively large margin of error.

Two potentially 'good enough' methods might be the use of evaluative rubrics[7] and multi-dimensional indicators. Both aggregate input data to a higher generic level, allowing for comparability across diverse types of data. **Evaluative rubrics** bring together different lines of evidence to assess performance based on a common set of criteria and standards, feeding into a transparent process of synthesising evidence into an overall evaluative judgement. Some potentially useful work has been done by OPM[8] and Julian King[9] on the use of rubrics for assessing VfM. **Multi-dimensional indicators** (usually indices) can be constructed to be more or less sensitive to variability of input data, and are useful for an assessment of relative performance – acknowledging that they can be simplistic and can disguise weaknesses in the source data. Examples include the ILO's productive employment indicator[10], the

---

[7] https://www.betterevaluation.org/en/evaluation-options/rubrics

[8] 'OPM's approach to assessing Value for Money' (2018)

[9] https://www.julianking.co.nz/vfi/4es/

[10] https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_565180.pdf

Grameen Foundation's Progress out of Poverty Index[11] and the UNDP Human Development Index. Such indicators are generally used to simplify a set of multiple indicators; but the method can also be used to smooth out some of the volatility of individual series.

Development and guidance on gender and VfM approaches

Across the evaluations, the approach to assessing gender and VfM has focused largely on reporting certain key metrics – number of female/male beneficiaries, extent of female engagement, administrative cost ratios, cost per beneficiary, etc. – but in most cases without analysing these to extract any very conclusive findings. In the case of gender, it would be good to see more in-depth and meaningful consideration of how programmes have engaged with gender issues than a numerical indicator of the extent to which women and girls are represented in some activities. In the case of VfM, while guidance and a broad '4Es' framework exist, with a number of evaluations drawing on similar metrics, this has not resulted in consistent approaches across the evaluations that might support synthesis, or in very strong conclusions.

It might therefore be practical for DFID to provide a stronger steer and guidance on what it expects in these two areas, providing an opportunity to coalesce around a key set of questions and issues. This is likely to require some preliminary analysis.

On gender, there is a substantial literature on the dynamics of women's participation and role across the three livelihood strategies and within an M4P context. This can be used to frame a common analytical approach tailored to the market systems framework, on the basis of which specific, relevant questions about gender can be articulated. This common approach would then provide a foundation for synthesis across different evaluations.

On VfM, evaluators may have sought to interpret the 4Es too literally: as Julian King points out, 'The Four Es … are not the last word on VfM criteria but they do offer a reasonable starting point … Sitting behind DFID's criteria are a more generalisable set of principles that can be applied more flexibly to respond to different contexts'. Importantly, they do not represent all of the possible dimensions of VfM. A VfM framework, based on the 4Es approach but more tailored to market systems programmes, might support better evidence and stronger conclusions that lend themselves better to synthesis. The blog post[12] and the OPM paper provide more detailed examples of how this might be applied.

[11] https://www.findevgateway.org/sites/default/files/publications/files/mfg-en-paper-poverty-targeting-and-measurement-tools-in-microfinance-progress-out-of-poverty-index-and-the-poverty-assessment-tool-oct-2010.pdf

[12] https://www.julianking.co.nz/vfi/criteria/

Annex 1. Summary of programme log frames

	FTESA	WAFM	PropCom	MADE	ELAN	LEAD	IMSAR	AECF
IMPACT								
	Improved national & regional staple food markets	Stable food prices in cross border markets	Sustained income increase for rural poor	Increased growth & income for poor	Reduced poverty, improved poor incomes	Enhanced economic wellbeing of farmers	Increased income & jobs	Growth of agriculture/ agribusiness which benefits farmers/rural poor
Indicator	stable prices	seasonal price differences	poverty ratio; av. hhd income	Net Additional Income for SMEs/farmers	poor farmer/ entrepreneur income	self-perception; access to food; borrowing	hhd incomes / job creation	Agricultural GDP growth % of people living on below $2/day
OUTCOMES								
Outcome 1	Increase in national & cross border trade in staples	Cross border staple food markets work better	Selected market systems work more effectively for poor	Better performing poor farmers & entrepreneurs in target markets	Well-functioning markets and business environment	Higher small holder incomes	Improved farmer/ entrepreneur access to markets	Agribusiness, finance and information market systems work better for the poor in rural areas
Outcome 2		More farmers benefit from national and X-border value chains	Improved resilience/ adaptive capacity for poor	Increased growth in 'aspirational sectors' for potential victims of trafficking			Innovation causing better functioning markets	Private sector investment catalysed, leading to development returns, innovation, and changes to market systems

	FTESA	WAFM	PropCom	MADE	ELAN	LEAD	IMSAR	AECF
OUTPUTS								
Output 1	Improved post-harvest markets	Policy facility to generate evidence, for policies better functioning cross border trade	Market systems work better for farmers & rural SMEs	Improved inputs, services, technologies introduced in target markets	Partner market actors have taken up pro-poor innovations and invest to sustain them [Adopt, Adapt]	Local and national markets accessed	Market actors operating in agricultural markets change practice and increase investment	Development benefits with high outreach, in particular to: those in rural areas; very poor; women; and youth
Output 2	Improved input markets	Improved business models developed for the production & marketing of staple foods across borders	CSA Practices adopted.	Market actors, inc. NGOs, Govt, change their approaches	Competing Market actors have copied the pro-poor changes [Expand]	The quality and availability of inputs and technologies will have been improved.	Partnership interventions brokered to address key constraints in agricultural markets	Business initiatives are sufficiently commercially viable that development impacts are sustainable
Output 3	Improved policy and regulatory environment		Market actors, inc. NGOs, Govt, change their approaches	Improved opportunities for (trafficking) susceptible populations in EDO & Delta States	Non-competing market players have adjusted to the pro-poor MSC. [Respond]	Access to agri-finance by smallholder farmers will have been improved		AECF identifies and supports business initiatives which generate sustained and significant positive development impacts and which are innovative
Output 4			High risk interventions for MSC			Demonstrable gaps in the value chains for maize and poultry have been addressed		AECF identifies, selects, contracts and provides managed support to business initiatives

	FTESA	WAFM	PropCom	MADE	ELAN	LEAD	IMSAR	AECF
INTERVENTION AREAS inferred from output indicators								
Input market	• Seeds • Fertiliser • Extension	• Inputs • Services	• Inputs • Services • Technology	• Inputs • Services • Technology	• 'pro poor innovations'	• Seed & feed • Technology • Agro-dealer training • Extension	• Inputs	
Output market	• Storage • Aggregation • Market Info • Grading	• Storage • Aggregation • Market Info			• 'pro poor innovations'	• Collective selling • Non-local sales • Contract sales	• Aggregation • Value addition	
Finance market	• WHS Receipts • Supplier Credit	• WHS Receipts • Supplier Credit • Mobile Wallets • Banking			• ? 'pro poor innovations'	• Group lending • New products		• Financial support to catalyse match funding and third party debt and equity
Policy reform	• Influencing Strategies	• Evidence • Influencing Strategies	• Regulatory change • Public funding					

Annex 2. Analysis of programme log frames

Impact statements and indicators

The two trade programmes are expected to result in improved national and regional markets for food. The stated indicator, stability of food prices within cross border food markets, reflects an implicit belief that the large seasonal price variations are a major factor for poor people.

The other six projects all show similar impact statements: that there will be a sustained increase in incomes or economic well-being for the poor. Some extend this to additional factors such as resilience. Impact indicators range from the standard household income and poverty ratios to broader livelihood indicators around access to food, jobs and borrowing.

Outcomes

The outcomes presented are quite general; and in some cases restate the impact in slightly different language. In others they are better considered as an Impact indicator. As such they do not make it easy to construct a strong logic and causal chain between outcome and impact, or to identify what contribution each outcome is expected to make to the impact.

Outputs

With very limited exceptions stated outputs are better considered as outcomes, as **changes the programme is expected to bring about**, as opposed to outputs which programme implementers are **directly responsible for delivering**.[13] Some examples:

- FTESA Output 1: Improved input markets
- PropCom Output 1: Market systems work better for farmers and rural SMEs
- ELAN Output 2: Competing market actors have copied the pro-poor changes [Expand]

Rather than outputs these are, in effect, intermediate outcomes leading into the higher level outcomes shown in each log frame. Given that those higher level outcomes are, in most cases, quite general, the log frames would be strengthened if they were removed to allow these outputs to be moved up to their proper place as outcomes.

There are cases where genuine outputs sit beside what are better described as outcomes in the same log frame:

- IMSAR Output 1: Market actors operating in agricultural markets change practice and increase investment (an outcome – a result of IMSAR's work)
- IMSAR Output 2: Partnership interventions brokered to address key constraints in agricultural markets (an output delivered by IMSAR)

This is not just log frame pedantry. Without a clear statement of what the programme will deliver as outputs, it is impossible to interrogate the most critical link in the ToC: between what the programme does and the changes it is expected to bring about. As well as weakening the ToCs, this also blocks any real understanding of value for money.

Intervention areas

Annex 1 shows an indicative summary of each programme's intervention areas. This has been inferred from the indicators shown against the outputs. In summary it shows that:

- All eight programmes are working in inputs markets to promote input supply, services and technologies, more broadly 'innovations'.

[13] This point is strongly noted in the CAPR.

- Six are working in output markets on storage, aggregation, market information, collective selling etc.
- Four are working in finance markets. Some financial interventions such as warehouse receipts (WHS) are linked to interventions in the output market.
- Three are working in policy reform, most especially the two trade programmes.

This breakdown comes directly from the log frames. In implementation, programmes may have extended the range of interventions to address different opportunities and challenges.

Annex 3. Summary of VfM indicators

FTESA	WAFM	ELAN*	LEAD	AECF	ProCom
ECONOMY					
Fund management cost ratio Administrative cost ratio		Daily personnel cost Total spend by project Total TA days on project Proportion of TA days performed by national team members	Overhead costs to total programme costs Ratio of facilitation costs (including training and direct service delivery) to overhead costs. (not reported) Ratio of training days delivered by external consultancy support to in-house staff. (not reported) Average fee costs for national and international consultants	Total expenses for fees and costs	*The ProCom VfM report was not provided for the review, but we have included this column to highlight the availability of additional data*
EFFICIENCY					
Smallholder engagement rate Leverage ratio Cost per MT capacity of warehouse construction/ refurbishment	Cost per farmer benefited Credit leverage ratio Cost per metric tonne of reported cross-border staples trade that the programme has contributed to	Proportional leverage of partners' investments Proportion of partners that sustain investments in an intervention Proportion of partners who have invested resources in an initial pro-poor innovation and intend to sustain it; Proportion of partners continuing activities which support the pro-poor innovation 12 months after the initial pilot has ended Cost efficiency by sector Cost efficiency by province	Cost per farmer adopting specific management techniques and technologies (no data) Cost per producer group formed and functional Leverage ratio Programme cost per £1 of private investment leveraged by the investment facility Programme cost per £1 of loan disbursed	Efficiency score USD spent per beneficiary household	

FTESA	WAFM	ELAN*	LEAD	AECF	ProCom
EFFECTIVENESS					
Volume of staple food sales per farmer reached		NAIC & discounted present value Proportion of poor people reporting substantial increase in enterprise or household performance who experience NAIC, by sector, province and productive/consumer saving Proportion of poor people reporting business practice change who also report substantial increase in enterprise or household performance, by sector, province and productive/consumer saving Proportion of poor people reporting business practice change who experience NAIC, by sector, province and productive/consumer saving Cost-effectiveness by sector & discounted present value Cost-effectiveness by province	Cost per farmer reporting a 10% increase in income (no data) Ratio of total programme cost to total net additional income above baseline reported by farmers Programme cost per poorer households Programme cost per household self-reporting improved access to sufficient basic food items Number of business starting operations in unserved areas Changes to marketing dividend for farmers	Net benefit recorded per USD spent	
EQUITY					
Percentage of women among smallholders engaged Cost of female outreach	Percentage of women among smallholders engaged				

* ELAN VfM metrics at the time of the MTE – prior to recommended revisions

Annex 4. Terms of reference

Evaluation Unit - Learning Review – Recent Agriculture Evaluations

1. Introduction

In February 2019, DFID reviewed its evaluation approach and determined that a combined centralised/decentralised system is needed to maximise opportunities for learning from DFID's own programmes and to strengthen the decentralised evaluation approach that DFID adopted in 2011. A key shift under the new approach is to focus upon conducting fewer but more strategic evaluations that make better use of the existing evidence base. To support implementation of this shift, Evaluation Unit are developing a new centrally managed programme to conduct thematic evaluations and mechanisms to support greater co-ordinating of evaluations at a thematic level, within priority sectors and/or geographies. The potential to promote greater use of consistent/comparable outcome and impact indicators within thematic areas as a means to draw together evidence and learning is also being explored. Initiatives are already underway within some departments within DFID to develop and use standardised indicators for core results. For example, DFID's Growth and Resilience Department are developing a set of voluntary core indicators for new DFID Commercial Agriculture programmes, following recommendations made by the second Commercial Agriculture Portfolio Review. [14]

Since April 2018, DFID has published (or is about to publish) several evaluations on commercial agriculture programmes in Africa. There are also evaluations of multi-donor, multi-country programmes (such as the Agriculture Enterprise Challenge Fund) covering similar areas. The availability of this body of evaluations provides an opportunity to draw together areas of common learning across these evaluations, as well as to consider how such evaluations might be adapted in future to enable more rigorous synthesis of findings and better learning on what works to reduce poverty through DFID agricultural programming.

This type of learning would be complementary to the findings and recommendations of the Commercial Agriculture Portfolio Review carried out for DFID in 2018-19 which focused on coherence of programming with shifts set out in DFID's Conceptual Framework on Agriculture (2015), collation and verification targets, results and budgets and analysis of specific aspects of results data. By drawing out analysis of learning on evaluation methods and indicators used across more than one intervention, this review should produce insights that can contribute to the progress already made by the Growth and Resilience Department towards developing a common set of core for DFID agricultural interventions in Africa.

2. Purpose and Objectives

The purpose of this learning review is two fold:

1) To draw together areas of common learning across these evaluations in relation to relevance to DFID policy and programme effectiveness, with a view to improving programme design and implementation.

[14] IMC Worldwide, Commercial Agriculture Portfolio Review 2018 for DFID, 2019.

2) To inform thinking on how to improve capability to evaluate impact and cost-effectiveness, and DFID's contribution through synthesis of learning from interventions with similar objectives and activities.

The findings of the review will be used by DFID staff designing and managing agriculture programming, evaluation advisors supporting evaluation of agriculture programming, DFID policy advisors and the Evaluation Unit. The review is intended to support DFID's internal learning and may be shared with selected external partners active in this field.

Review questions

- What are the areas of common learning across the evaluations?
  - For the ARD multi-country evaluations - do the evaluations identify any specific advantages or challenges involved with taking a regional or multi-country approach to provide support in this area?
- To what extent are the outcome and impact statements of the programmes examined coherent with the directions set out in DFID's Conceptual Framework on Agriculture;
- What were the advantages/disadvantages of the evaluation methods used? Would other forms on evidence generation provide more effective ways to meet evidence gaps?
- Where the evaluations examined common questions or TOC assumptions, to what extent do they provide consistent evidence for or against these?
- Were there any opportunities to build in practical measures to make it easier to synthesise learning across programmes in future?
  - For example, are there any common evaluation questions, or common metrics, or common VFM measures, that have been used successfully across more than one of the evaluations reviewed?

3. Scope

The learning review should cover the agriculture-related programmes listed below in Table 1 which have all undertaken evaluations that have been published since April 2018 or are about to be published and those in Table 2 which have not commissioned evaluations but have used alternative approaches to gather necessary evidence and learning.

Table 1 Programmes with Evaluations

Programme	Department	Type of evaluation	Programme Start Date	Programme End Date	Publication date
FoodTrade East and Southern Africa (FTESA)	Africa Regional Department	Mid-term	2013	2018	2018/19
FoodTrade East and Southern Africa (FTESA)	Africa Regional Department	Final	2013	2018	Not yet published
Rural and Agriculture Markets Development programme for Northern Nigeria (PrOpCom Mai-karfi)	DFID Nigeria	Final	2013	2020	2018/19

Market Development in the Niger Delta (MADE)	DFID Nigeria	Final (Independent Review not Evaluation)	2014	2020	Not yet published, Phase I report available end Aug 2019
West Africa Food Markets	Africa Regional Department	Mid-term	2013	2019	Not yet published
ELAN programme (agriculture-related components), Private Sector Development Programme	DFID DRC	Mid-term	2012	2024	2018/19
Livelihoods Enhancement through Agricultural Development	DFID Tanzania	Mid-term	2013	2017	2017/18
Strengthening Indian Trade and Investment for Africa (SITA)	DFID Ethiopia, Kenya, Rwanda, Uganda, Ethiopia	Mid-term	2014	2022	2018/19

Table 2 – Programmes using alternative approaches to evidence and learning

Programme	**Department**	**Programme Start Date**	**Programme End Date**	**Evaluation approach**
Improving Market Systems for Agriculture in Rwanda (IMSAR)	DFID Rwanda	2015	2022	4 MEL reports per year (3 quarterly MEL reports plus annual evaluation report), being amended to 2 MEL products per year
Northern Uganda: Transforming the Economy through Climate Smart Agribusiness (NU-TEC)	DFID Uganda	2014	2022	Independent monitoring with additional learning products.

The learning review should also examine relevant sections of programme documentation for these programmes – Business Case, annual reviews, programme completion reviews (PCR) and results frameworks.

The learning review should also take into consideration three evalations/reviews of multi-country, multi-donor commercial agriculture programmes in which DFID has been a key stakeholder, listed below:

Table 3 – Multi-donor commercial agriculture evaluations and reviews

Programme	Type of Evaluation	Publication Date
Africa Enterprise Challenge Fund	Mid Term Evaluation (and Final Evaluation, if available in time)	August 2015
Global Agriculture and Food Security Programme	Final Evaluation	Jun 2018
Enabling the Business of Agriculture	External Review	May 2018

The study should be carried out as a desk review of the available documents and by conducting interviews by phone or skype with SROs and/or evaluation advisors for the programmes included to validate and refine initial findings from this learning review.

4. **Deliverables**
   i. Work plan - the reviewer should produce an initial work plan outlining the proposed approach and analysis framework for the review and time line for completion.
   ii. Report – the reviewer should produce a concise, logically structured report that meets the overall purpose and objectives of the learning review and provides responses to the questions set out in section 2. A draft version of the report should be shared with DFID for feedback and a final version should be submitted addressing feedback received.

5. **Skills and experience required**

- Significant experience and expertise in agriculture and management of agricultural development programmes in Africa
- Significant evaluation experience, qualitative research and analysis skills
- Excellent communications skills and ability to distil succinct conclusions presented in non-technical language.
- Familiarity with DFID programming systems and monitoring/review processes.

6. **Timeframe**

This work should take place between August – October 2019.

7. **Time allocated**

A total of 25 consultancy days is allocated for completion of this assignment.

8. **DFID Coordination**

The point of contact in Evaluation Unit for this work will be Louise Davis, Evaluation Advisor (Thematic Lead). She will co-ordinate engagement and feedback from key stakeholders in the review within DFID including

SROs and Evaluation Advisors in Country Offices, Africa Regional Department and Economic Development Growth and Resilience Policy Team.

9. Other requirements

- Compliance with DFID's Ethics Principles for Research and Evaluation
- DFID will have unlimited access to the material produced by the supplier in accordance with DFID's policy on open access to data/as expressed in DFID's general conditions of contract

10. Background

DFID's work on Commercial Agriculture is an important component of DFID's overall portfolio to support economic development. Its primary target group are smallholder farmers who are 'stepping up' and engaging in agriculture as a commercial activity, rather than 'hanging in' as subsistence farmers. It should be noted that some of the projects and programmes which include commercial agriculture within their scope are multi component or multi-sectoral and not all of the commitment is to commercial agriculture.

Thus the Commercial Agriculture Portfolio Review was undertaken in 2017 and 2018 to provide an up to date compilation and analysis of the DFID's programmes on commercial agriculture. The Portfolio Review examines key areas such as alignment with strategy, resource allocation, monitoring, reporting and verification systems and indicators for results and achievements and to inform future programming.

A key recommendation from the CAPR is greater use of standardised indicators for core results to enable DFID to obtain a clearer picture of the impact of our work, inform programme decision-making and increase accountability with robustly measured results that can be aggregated. The proposal to develop a set of standardised indicators for voluntary use has been approved by the Director General for Economic Development and International (Nick Dyer). Thus, this review could also generate learning that would support the development and use of more consistent and comparable indicators for DFID agricultural interventions in Africa.