



Towards better evidence for market systems initiatives

Think piece

Elise Wach

December 2015

Executive Summary

Debates about the need for rigorous evidence are increasingly common in international aid and development circles. This paper asks what constitutes credible evidence for programmes and initiatives that use a market systems approach.

Calls for rigour are often misinterpreted as calls for certain methods which are deemed as rigorous. While methods are important, this piece argues that what is needed for the generation of rigorous evidence goes far beyond a question of methods, and specifically, beyond measuring progress against predetermined targets.

Pragmatically, not all studies need to be rigorous to be valuable and not all programmes are of a scope to merit investments in rigorous assessments. However for those that do, this think piece aims to strengthen understandings about what constitutes rigour, and suggest potential ways in which improved evidence could be generated.¹ It draws on consultations with individuals in the field of market systems development and a review of literature related more broadly to evidence.

Initiatives using a market systems development approach² seek to achieve changes in markets which are both large-scale and enduring in the long-term. These changes in market systems are in turn expected to achieve positive outcomes for people living in poverty; namely poor people who are (or have the potential to be) producers, consumers or employees in a given market ('beneficiaries'). Key features of the market systems development approach include its indirect facilitative nature, its adaptive approach and the fact that it often does not target the poorest of the poor but rather aims to positively impact these people through 'multiplier effects' or through growth. The approach also draws on systems thinking concepts and to some extent on the premise that markets are complex.

A fundamental principle in assessments is that methods must align with the assessment question and the context. For market systems initiatives, it is broadly agreed that experimental methods are often a poor fit given the adaptive nature of the approach. It is also arguable that qualitative approaches have great potential for linking interventions with outcomes and impacts, but need to be applied in a more rigorous way.

In addition, given the unique features of the market systems development approach, the scope of assessments, i.e. what is assessed and not assessed, is particularly relevant for generating rigorous evidence. Changes in the systems of the poor, just like changes in market systems, are nonlinear and unpredictable. An increase in income for one segment of the population, for example, cannot be assumed to be the only significant effect which has occurred as a result of a market system change.

Based on principles of systems thinking and complexity theories, in order to be rigorous, assessments of market systems initiatives should generate an understanding of the multidimensional changes which have occurred as a result of changes in market systems (Befani et al, 2015). This is particularly important for changes which are experienced by people living in poverty, even when such people are not intended as beneficiaries. In

¹ This piece does not attempt to provide a comprehensive review of existing evidence or approaches to generating evidence

² Referred to in this paper as 'market systems initiatives' for ease

practice, complete understandings are impossible but market systems initiatives can move towards greater understandings and therefore greater rigour by incorporating a plurality of perspectives in assessments and by privileging the perspectives of people living in poverty. In addition, utilising approaches which are congruent with complexity can enable assessments to capture the broader ‘emergent’ changes associated with an intervention that might not have been anticipated by programme implementers or evaluators, but which may be important. Incorporating these elements in existing efforts to assess market systems initiatives could increase the rigour of evidence available about these initiatives by providing a more accurate understanding of their effects and significance for people living in poverty.

In addition to assessment approaches themselves, the existence of strong ‘learning cultures’ is crucial for generating rigorous evidence. Practitioners consulted for this piece indicated that more openness about failure needs to exist both internally (i.e. among implementing and evaluating staff), as well as externally (i.e. between programme teams and donors). Without this openness, rigorous evidence about the full effects of a programme may not be generated or acted upon.

Introduction

Initiatives using a market systems development approach³ seek to achieve changes in markets which are both large-scale and sustained in the long-term. These changes in market systems are in turn expected to achieve positive outcomes for people living in poverty; namely poor people who are (or have the potential to be) producers, consumers or employees in a given market (‘beneficiaries’).

In the context of increasing demands for rigorous evidence for international aid and development initiatives (Du Toit, 2012), there have been some debates as to the extent to which rigorous evidence exists about the effects of market systems initiatives. These debates raise questions about what constitutes rigorous and credible evidence in the case of initiatives (or programmes) that use a market systems development approach, and how much of it is needed in order to adequately inform decision making.

This think piece explores these questions, drawing on consultations⁴ with practitioners, donors and observers of market systems initiatives, as well as wider literature about evidence and rigour from beyond the fields of market systems and of development. Rather than providing a comprehensive review of existing approaches or evidence, the piece aims to strengthen understandings about what constitutes rigorous evidence for initiatives implementing market systems approaches and suggests potential ways in which improved evidence could be generated.

Mainstream debates about rigorous evidence for market systems initiatives include, amongst other issues⁵, those relating to methods, the extent to which changes are ‘rigorously’

³ Referred to in this paper as ‘market systems initiatives’ for ease

⁴ Consultations included interviews with ten practitioners, donors and observers of market systems initiatives and an [online discussion](#) on this topic with 18 active contributors and 108 registered participants.

⁵ Rather than comprehensively capturing all of the issues related to this topic, this section focuses only on a selection given the limited scope of this paper.

measurable and the importance of organisational culture in the generation of (rigorous) evidence.

In addition, this paper argues that other issues need to be included in the debates about evidence for market systems initiatives. Namely, it argues that in order to be rigorous, evaluations and assessments which are congruent with systems thinking and complexity theories would go beyond the use of targets and indicators to assess the effects (foreseen and unforeseen) for both intended beneficiaries and for other segments of the population (Befani et al., 2015; Williams, 2015). This paper suggests that one way in which assessments can move towards this scenario is by incorporating a greater plurality of perspectives, including those of non-beneficiaries, non-stakeholders and other ‘unusual’ suspects—in addition to those of beneficiaries and stakeholders—in order to generate a more robust understanding of a programme’s effects and the significance of those effects for different people.

Pragmatically, not all studies need to be rigorous to be valuable and not all programmes are of a scope or significance to merit investments in rigorous assessments. However for those that do, this think piece aims to strengthen understandings and suggest potential ways in which improved evidence could be generated.

The piece is structured as follows. The first section provides a brief background and overview of some of the mainstream debates. This is followed by discussion of a selection of wider issues related to the rigour of evidence which go beyond what dominates the mainstream debates. The subsequent section presents a synthesis of guiding principles for generating rigorous evidence in the context of market systems interventions. The paper ends with a short summary and concluding remarks.

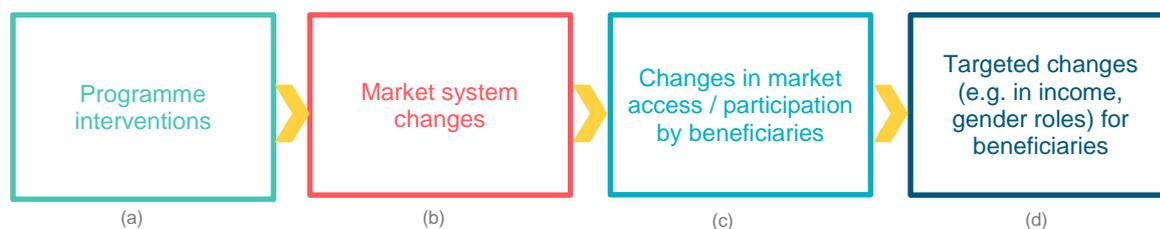
1. Background and highlights of the mainstream debates

a. Background

The premise of market systems development approaches is that they aim to achieve enduring or lasting (often referred to as ‘sustainable’) changes at large scales through facilitating changes in market systems. It is expected that changes in market systems in turn result in changes in the ways in which beneficiaries engage in or are affected by markets (e.g. as consumers, producers or employees). This in turn is expected to lead to changes in development outcomes for beneficiaries (e.g. increases in production, increases in income, improved working conditions, access to services, etc.).

Efforts to generate evidence about the effects of market systems initiatives thus include a focus on each of these steps. Broadly, this includes monitoring the interventions, assessing changes in market systems, assessing changes in beneficiary participation in market systems, and assessing (or estimating) changes in the outcomes for beneficiaries. Figure 1 provides a simplified representation of this logic.

Figure 1: Simplified logic of market systems development approaches



In order to achieve this in practice, many initiatives using a market systems approach have designed their Monitoring and Evaluation (M&E) or Monitoring and Results Measurement (MRM) systems based on guidance from the Donor Committee for Enterprise Development (DCED). The DCED Standard for Results Measurement⁶ was developed for initiatives in the field of private sector development (PSD), which includes but is not limited to market systems development approaches. Projects and programmes implement the DCED Standard by following guidelines for ‘(1) articulating the results chain, (2) defining indicators of change, (3) measuring changes in indicators, (4) estimating attributable changes, (5) capturing wider change in the system or market, (6) tracking programme costs, (7) reporting results, and (8) managing the system for results measurement’ (DCED, 2015). According to the DCED website, more than 100 projects are currently using the DCED Standard. A number of people consulted for this piece referred to the usefulness of the standard in helping projects to establish an MRM system for generating evidence and informing decision making.

Many practitioners indicate that the DCED Standard has improved the quality of evidence generated for market systems initiatives. This paper aims to build on the progress of the DCED Standard in order to further improve the rigour of evidence generated about market systems initiatives.

b. Which methods are rigorous?

In initiating this review, it was anticipated that debates about evidence would be dominated by discussions of methods. ‘Is a randomised control trial (RCT) a gold standard?’; ‘is an experimental design inappropriate for market systems?’; ‘should we try SenseMaker®?’ for example. While methods are relevant to the debates about evidence, it is widely acknowledged⁷ in the broader literature about evidence (i.e. those outside of the field of development) that there is no absolute hierarchy in the rigour of different methods: a method which might be rigorous in one context can be completely irrelevant in another (Nutley et al, 2013; Stern et al, 2012, Sackett and Wennberg, 1997, Pawson, 2003). A DFID working paper on ‘broadening the range of designs and methods for impact evaluation’ unequivocally states, ‘The choice of methods (and overall designs) needs to follow from the kinds of questions that are being asked; whilst also taking into account the settings in which they are to be applied.’ (Stern et al 2012: 9).

⁶ Commonly referred to as the ‘DCED Standard’. More on the DCED Standard for Results Measurement can be found at <http://www.enterprise-development.org/page/measuring-and-reporting-results>

⁷ Despite wide acknowledgement of fitting methods with questions, there are undoubtedly some people who still advocate that certain approaches (whether RCTs or qualitative studies) as absolutely better than others.

All of the practitioners and individuals working within donor agencies consulted for this piece recognised that the method needs to match both the question and the programme context, thus departing from the concept of an absolute hierarchy of methods. As one respondent stated, ‘we need to not advocate a particular methodology: that’s where the problems come in.’

There was, however, general consensus that experimental methods (e.g. RCTs) typically do not work for market systems initiatives, particularly because of the adaptive nature of programme designs. One practitioner stated, ‘One thing that doesn’t work is to rely heavily on large surveys...this can cause a big problem. People will manoeuvre themselves into a corner, and end up empty handed.’ An individual associated with a donor agency stated that ‘Because of the adaptive nature it is difficult to hold the programme to results against a baseline. It makes experimental approaches very difficult.’ Another practitioner pointed out, ‘It’s not the type of project that you plan in the first month and then execute for the next three years. It’s adaptive.’

While experimental methods were rather largely dismissed as inappropriate for market systems interventions, there was some disagreement about what methods *would* produce rigorous and credible evidence. Some called for more quasi-experimental⁸ studies, indicating that any other approaches would produce anecdotal and unreliable estimates.

Some people equated qualitative approaches with ‘glossy’ or ‘fluffy’ case studies, of which there is a perceived abundance. While glossy case studies are certainly not rigorous, there is great potential for qualitative evaluation methods to be conducted with rigour. In fact, they can often link interventions with outcomes and impacts in a way that experimental and quasi-experimental approaches cannot (Stern et al., 2012). As Michael Quinn Patton states⁹, ‘Cherry-picked anecdotes to supposedly “prove” a predetermined position come across as what they are: argumentative advocacy, not evidence. But the systematic, intentional and careful recording of purposefully sampled anecdotes (stories) can become evidence when rigorously captured and thoughtfully analyzed’ (Patton, 2015).

The opinion that qualitative approaches are inherently ‘fluffy’ or that their main purpose is to ‘support’ quantitative assessments could reflect the way in which qualitative studies (often case studies) are typically conducted in market systems evaluations (i.e. generally light touch with limited rigour in the selection of participants, data collection and analysis (Ruffer and Wach, 2013)), as much as it could reflect an uninformed bias towards (or against) particular methodologies. Thus, there remains work to be done to improve both the conduct of

“Cherry-picked anecdotes to supposedly “prove” a predetermined position come across as what they are: argumentative advocacy, not evidence. But the systematic, intentional and careful recording of purposefully sampled anecdotes (stories) can become evidence when rigorously captured and thoughtfully analysed” (Patton, 2015).

⁸ For an overview of the difference between experimental approaches (e.g. RCTs) and quasi-experimental approaches (e.g. difference in difference, regression discontinuity studies, propensity score matching), see http://www.sagepub.com/upm-data/33935_Chapter13.pdf

⁹ http://betterevaluation.org/blog/anecdote_as_epithet

qualitative assessments, which could help, in turn, to address any potential biases.

In sum, certain methods are not necessarily more rigorous than others; what matters is the fit of the method with the question and the context. In the context of market systems initiatives, it was broadly agreed that experimental methods are often a poor fit for the adaptive nature of programmes, and that qualitative approaches could be applied in a more rigorous way. Later in this piece, the issue of the assessment questions will be further explored.

c. What is ‘rigorously’ measurable in the context of complexity?

Those implementing or advocating for market systems development approach indicate that what is unique about the approach is its indirect nature (Springfield Centre, 2014). Changed outcomes for people living in poverty are expected to result from changes in market systems, rather than from direct engagement with the programme activities themselves. Examples of other relatively indirect approaches include advocacy and policy change programmes (Jones, 2011).

In addition to the indirect nature of the approach, many people working within the field of market systems development often point to the fact that because markets are complex, it is difficult to assess changes in them and also to attribute those changes to interventions. Specifically, the high number of interrelated actors and institutions operating in markets means that change does not occur in a linear way and can often be unpredictable. In addition, changes can take various amounts of time to occur – some changes may happen rapidly while others may take years to manifest. These challenges are experienced for relatively indirect initiatives in general, such as social policies and programmes (Levitt, 2013) and have been recognised in the case of market systems initiatives in particular (Ruffer and Wach, 2013).

One of the online discussants provided an example to illustrate the difficulty in assessing and attributing changes. The participant described an initiative which helped to legalise raw dairy traders, which in turn had a number of effects for both the traders and the suppliers. The participant stressed the difficulty in generating evidence to demonstrate how the different (and evolving) activities that helped facilitate the legalisation process (e.g. training of trainers for certification, etc.) could be linked to the multitude of changes that materialized in the sector.

On the other hand, some practitioners and donors interviewed perceived the talk of complexity as a crutch or an excuse for not demonstrating programme results. These practitioners indicated that there is a great deal (more) that programmes can in fact assess, despite the complexity of markets. In an attempt to tackle the issue of assessing systemic change, many programmes have started using the Adopt-Adapt-Expand-Respond (AAER) framework (Nippard et al., 2014). This framework attempts to capture signs that innovations have been ‘adopted’ and ‘adapted’ by companies, that other businesses have copied innovations (‘expansion’) and that other market actors or functions have changed in ‘response’ to the market intervention. The framework attempts to mitigate, at the level of markets (point b in Figure 1), the challenges in assessing complex changes and is included in the DCED Standard guidelines for assessing market systems change (Kessler, 2014).

At the level of beneficiaries, however, several practitioners argued that unless effect could be ‘rigorously’ measured, it was not worth attempting to assess. For example, one practitioner, in discussing an agricultural markets programme, reasoned that it was not possible to

rigorously assess any changes in beneficiaries beyond differences in production (point (c) in Figure 1) and therefore, this was the primary focus of assessments in the programme. His argument is that any claims about changes to poverty levels or even income (point (d) in Figure 1) would be conjectures and therefore unreliable.

‘All I can do is [measure an] increase production...the tonnes of rice, but not income or poverty...if you go above this, you get lies and hypocrisy...you can calculate average increase in income but it is based on assumption and often there is displacement.’

Other practitioners also shared this belief that ‘considering any effects beyond [productivity] is impossible to rigorously prove’ and therefore, assessments should not attempt to capture them. The practice of calculating and inferring (rather than directly assessing) changes in profitability for farmers (and often changes in income) based on changes in production level is nearly universal in market systems initiatives related to agriculture.

However, by **only assessing that which is perceived to be measurable in a ‘rigorous’ way**, it is possible for programmes to miss out on important evidence about the effects of their interventions. This has been shown to lead to misleading conclusions in other fields such as medicine and social policy. In a review of evidence related to the efficacy of a social policy, Pawson (2003:24-27) demonstrates how a prospective simulation, which by some standards (e.g. those based on a hierarchy of methods) would have been unlikely to be considered as ‘rigorous’ proved to be a high quality and valuable study. It provided deeper insights into the effect of the policy than those provided by a quasi-experimental study or an open-ended survey on the same issue. This is because the prospective simulation included an **explanation of patterns** (rather than simply showing a change or non-change), which was well-suited to the complexity of the policy and social environment assessed. As mentioned in the above section, methods must be fit for purpose in order to produce rigorous evidence. Occasionally, the best fit method—i.e. the one that will provide the most valuable insights—may not align with what is *conventionally* thought of as rigorous.

d. The cultures needed for generating (rigorous) evidence

In discussing the production of evidence, the need for a learning culture, which acknowledges failure as a natural part of a learning process, was recognized by a number of practitioners consulted for this think piece. It was stated that without such a culture, it is not possible for rigorous evidence to be generated and/or utilised in decision making.

Interviewees noted that the fear of discovering failure can prevent implementing teams from generating (or communicating) evidence about a project’s impacts. Specifically, practitioners pointed to the need for a learning culture to exist both internally and externally.

One team leader discussed how the lack of openness about failure prevented a previous programme from generating (and acting on) credible evidence about what was really happening as a result of the programme. He therefore strived to cultivate a different culture in his next endeavour:

‘In [the previous programme] our staff did not tell us what was happening and there was enormous cover up and selective information...We need a culture in which people can say ‘it’s not working’ and also ‘why it isn’t working’, both reflecting on their own work and that of their colleagues... It is very easy to be selective and you will be

if you don't have this culture. In [our current programme] we have been 80% successful in this. We are constantly praising [learning from] failure and giving feedback. In [the other programme], we didn't do that.' –M4P team leader

This individual noted that the programme management team needs to encourage and welcome the acknowledgement of failure. Specifically, it is important for staff to feel safe in acknowledging and communicating about what is not working.

Another team leader discussed how he tries to do this in all of his work:

'For me, it's stupid to assume that things work, so I try to ingrain in the people I work with that mistakes will happen. People assume that programme success is tied to their career, but that isn't the case.' –M4P team leader

Externally, donors and the public to which they are accountable need to be aware of failure as not only a possibility but as an integral part of any attempt at intervention. At present, communicating mistakes or failure to a donor is perceived as requiring a great deal of courage:

"It takes courage from management to go back to the donor and say, 'you know that thing that we said was working, actually we found out that it doesn't work'" –M4P team leader

"People probably know that what they did didn't make a big difference...there are some people with high integrity, who can be honest, but under pressure people make things look better than they are." –M4P team leader

The M4P Operational Guide, which is used by many market systems initiatives, also stresses the importance of a learning culture and attitude: 'Of equal importance to MRM system design is staff attitude. Management has to develop a culture that is learning-focused and that endorses self-critique. It is this that makes adjusting interventions possible. This has implications for how your MRM system is set up, staffed and resourced. The system and culture together is what allows programmes to respond to complex environments' (Springfield Centre, 2014:40).

Despite increasing recognition of the need for learning (including revisiting the underlying programme assumptions), it is still difficult for many to put into practice, in part due to the time and resource demands experienced by practitioners (and arguably donors). One practitioner involved in an M4P programme stated:

"[The team] realised that [this intervention] was not necessarily what is needed to address these issues, but we already had our [plans] approved and after spending the money on the inception phase we had to move forward with it." –M4P M&E officer

In considerations of how to produce credible, rigorous evidence for market systems initiatives, it is important to continue to work towards cultivating the required culture and leadership as much as it is to focus on the methods and approaches.

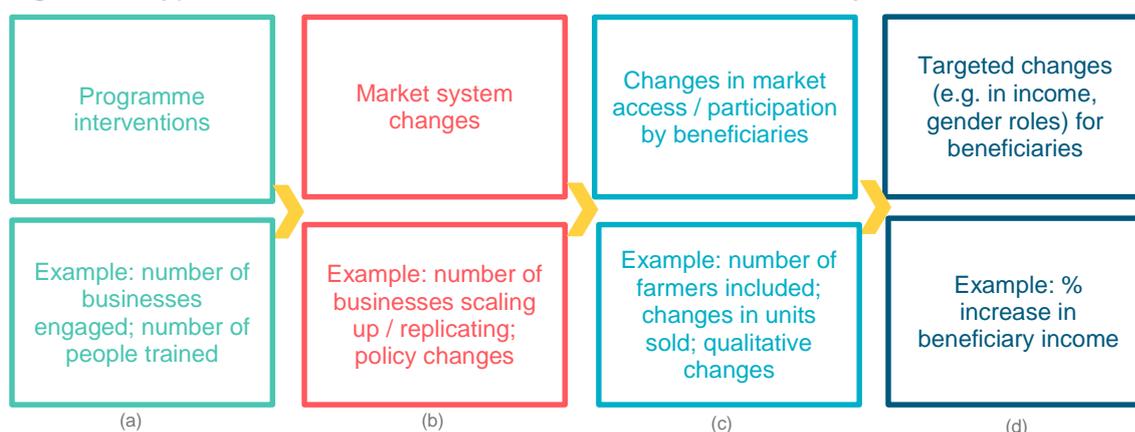
2. Issues missing from the mainstream debates

The concept of rigour does not only relate to whether a specific effect has been measured in a reliable and objective way. A number of factors determine whether a study is rigorous, including asking the right questions (i.e. the scope of the assessment), appropriate design (i.e. method fit for purpose), trustworthy data, verification, utility (appropriate to the needs of those affected by it), transparency, and accessibility to name a few (Morse and Richards, 2002; Pawson et al., 2003). This section will specifically explore the scope of assessments, including the processes which determine what is assessed and the people and perspectives which are included in the assessments.

a. Beyond targets to assess changes in the multidimensional systems of the poor

While assessments of market systems initiatives at the level of markets often incorporate systems thinking and complexity perspectives, approaches to assessing changes experienced by beneficiaries (points (c) and (d) in Figure 2) often follow a fairly simple and linear logic. They include assessing changes in production in the case of agricultural programmes (which make up the majority of M4P initiatives) and changes in employment in the case of other producer or employee focused approaches. In the case of consumer-focused programmes such as those aiming to increase the availability of certain services or products (e.g. micronutrient powders) to people living in poverty, outcomes measured typically focus on the number of products distributed (Sarwar, 2015). Some programmes may attempt to assess changes in income experienced by beneficiaries, though changes in income is often calculated by programme teams or evaluators based on changes in production rather than directly assessed through obtaining feedback from households¹⁰.

Figure 2: Typical areas of focus in assessments of market systems initiatives



Exclusively assessing changes in productivity, incomes, or access to a product or service carries the risk of simplifying poverty reduction to a linear process i.e. ‘identifying very specific, narrowly defined outcomes to be achieved, and then looking for evidence that can

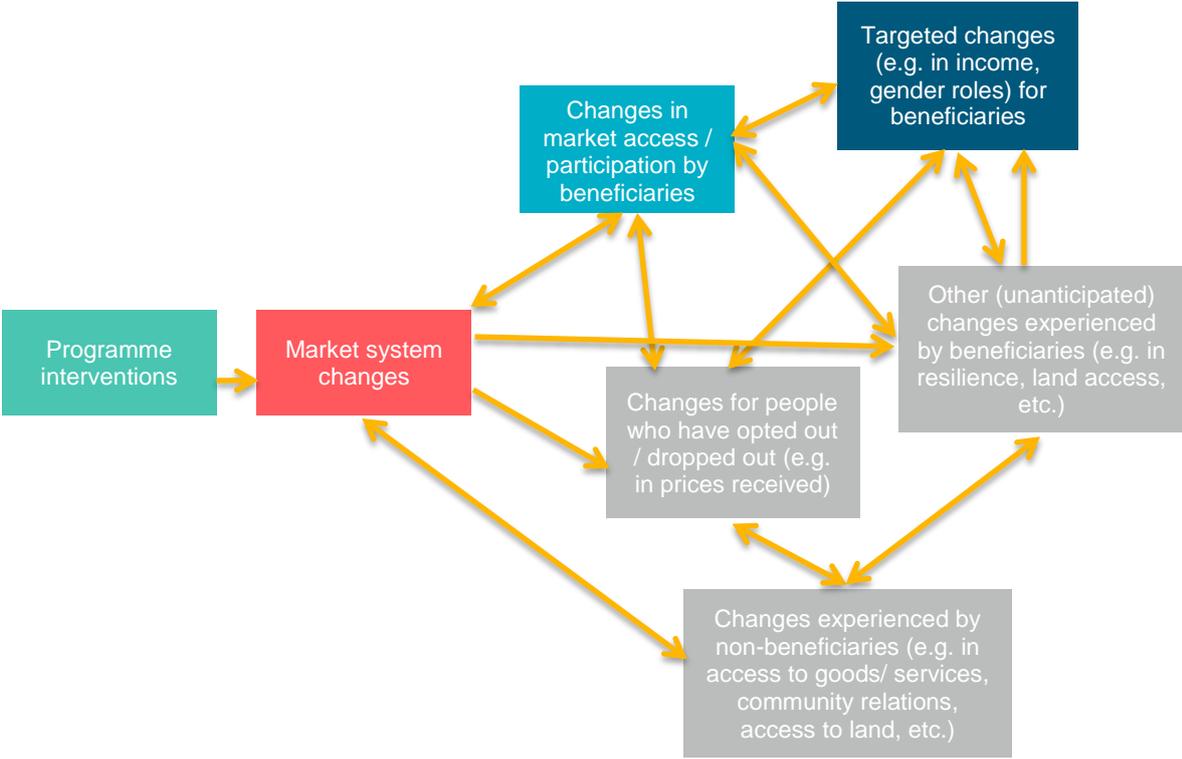
¹⁰ This practice of calculating changes in income rather than directly assessing it was noted as common practice by interviewees and has also been observed in the author’s experience with market systems development programmes.

allow clear and unambiguous attribution of these outcomes to distinct and clearly identifiable interventions’ (Du Toit, 2012:3).

For example, most studies on the effects of contract farming (not just those related to market systems approaches) focus on whether contract farming increases incomes for participating farmers. Studies may even go as far as to look at which types of farmers (e.g. moderately wealthy farmers, moderately poor farmers, etc.) are able to participate. However, most assessments of contract farming overlook the implications for non-contract farmers and excluded labourers and implications for the local economy (e.g. higher food prices for rural households, changes in land access) (Vicol, 2015 drawing on Oya, 2012 and Fold, 2009).

Taking into account the complex social, cultural, economic and environmental systems within which poor people live requires a non-linear approach. A more accurate depiction of changes which might take place as a result of changes in market systems might resemble the situation in Figure 3, in which changes – both anticipated and unanticipated – for different groups of people can be interrelated.

Figure 3: Multidimensional and non-linear changes which can result from market changes



Only measuring what a programme sets out to achieve is a significant but often overlooked potential source of bias in research and evaluation (Chambers, 2015) and is particularly problematic if evaluation efforts are to draw on systems thinking and complexity concepts (Williams, 2015; Garcia and Zazueta, 2015). To illustrate why, imagine a pharmaceutical drug trial which very rigorously assessed the ability of the medication to treat the target ailment but did not assess any other effects that the individual might experience, or any population-wide effects (e.g. antibiotic resistance) that could occur. It is difficult to argue that such a trial could be considered as providing an unbiased assessment of the effects of the drug in question. Current practice in medicine defines ‘quality evidence’ as ‘the amount of

confidence that a clinician may have that an estimate of effect from research evidence is in fact correct for both beneficial and harmful outcomes' (Nutley et al., 2013: 11, drawing on Atkins et al., 2004).

A few people interviewed for this piece voiced concerns about the narrow focus on targets in assessments of market systems initiatives. Another respondent discussed the concept of 'construct validity' in relation to assessments:

"What are we really measuring? We may need to pay more attention to construct validity. How good are these indicators?" –M4P evaluator

Applying concepts of 'construct validity' to market systems initiatives, as this individual suggested, would require assessments to consider whether our 'constructs' of 'success' or 'effectiveness' are valid. Again, going back to the medical trial, if the drug is found to reduce a particular symptom and is considered 'effective' but also causes strong negative side effects, then 'effective' has been defined as only related to the target condition. From a 'construct validity' perspective, that particular construct of 'effective' is arguably not valid¹¹ (Cook and Campbell, 1979).

In plain terms, it might not matter if a programme is able to rigorously assess a given change if that change is too narrow, and a study will not be rigorous if it does not capture all of the significant effects of the programme (Wach, 2012). A recent set of articles on systemic approaches to evaluation (Befani et al., 2015) **unequivocally called for evaluations to go beyond targets** in order to provide a rigorous (systems based) assessment of the effects of an intervention.

According to Stern et al 2012, 'evaluations must go beyond 'did it work' to include...equity questions, 'for whom do interventions make a difference?' (2012:11). This is particularly relevant for market systems initiatives given that **almost all of the individuals consulted for this piece recognised that market systems programmes often cannot and should not be targeting the poorest**. Rather, it is the less poor, 'emergent' individuals who are often able to take advantage of market opportunities, and who are able to be entrepreneurial, who benefit. A predominant (though not universal) theory of change is that the poorest will benefit from 'multiplier effects'.

"Should we really work with the poorest of the poor farmers or with emerging farmers and expect spillover effects? If we work with the emergent farmers, either they generate employment for the poor farmers, or at least the poor farmers will copy what these farmers are doing." – Individual working for an M4P donor

"Twelve years ago the general opinion was that programmes like Katalyst had to aim for growth. The assumption was that the poorest would also benefit. Now in M4P we are meant to start with the poor, but I personally don't believe in that...the poorest are never the innovators, they cannot take risks. They will have to benefit from growth which takes place" – M4P team leader

¹¹ Technically this is referred to as 'restricted generalisability across constructs' (note this is '*constructs*' and not '*contexts*')

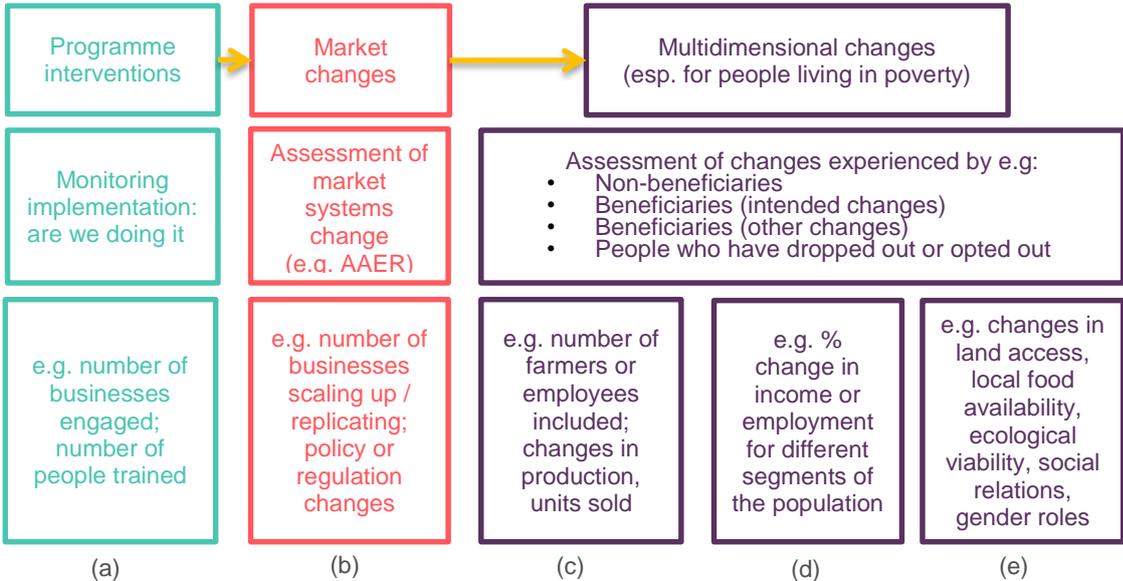
In current practice, assumptions underpinning these theories about how poorer populations will indirectly benefit (e.g. from local growth) often remain untested. While some assessments might consider which types of people are benefitting from an intervention, it is rare that an assessment would look at the effects experienced by the extreme poor¹², particularly when they are not explicitly targeted or intended as beneficiaries.

If assessments focus only on the changes experienced by non-poor or moderately poor target beneficiaries, and changes experienced by companies or other market actors, it is not possible for programme teams or evaluators to gain a credible understanding of the changes experienced by people living in poverty, especially when they are not intended as target beneficiaries.

Assessments which include not only the beneficiaries but also individuals *not* targeted by the initiative, particularly those living in poverty, are essential if we want to have rigorous, credible evidence about the effects of market systems initiatives on poverty. This would bring market systems initiatives into better alignment with Organisation for Economic Cooperation and Development (OECD) Development Assistance Committee (DAC) standards for evaluation which include ‘positive and negative, primary and secondary long-term effects on final beneficiaries; direct and indirect causal contribution claims of these interventions to such effects *especially for the poor whether intended or unintended*’ (Stern et al 2012; OECD-DAC 2010, emphasis added).

Figure 4 indicates how the typical framework for assessments of market systems initiatives might be revised to incorporate changes beyond the targeted effects for the beneficiary population, and therefore provide more rigorous evidence about their effects.

Figure 4: Proposed framework for more rigorous assessments of market systems initiatives



¹² Two discussants contested the concept that programmes do not target or assess their impacts on the extreme poor

Ongoing assessments of the multidimensional systems of people living in poverty could maximise the efficacy of programmes while increasing the rigour of evidence about their impacts. Multidimensional here means considering social relations, gender roles, ecological viability, food security, health or other factors which are deemed to be significant by the people experiencing these changes. Just as ongoing assessments of market systems can enable programme managers to adapt to changing market conditions and demonstrate lasting impact, understanding multidimensional changes can enable programmes to better respond to these complex systems as they change and as more insights become available about them.

b. Incorporating a plurality of perspectives

The issue of ‘perspectives’, while discussed by some participants, often does not feature prominently in the mainstream debates about evidence for market systems projects. However, it is arguably an essential component of evaluations which are congruent with systems concepts: ‘Systems-influenced evaluation deliberately exposes our (evaluators’ and stakeholders’) assumptions about what is valid knowledge. This exposure comes through our embrace of multiple perspectives, a critical systems concept’ (Imam et al., 2006:9)

In a recent publication about systemic approaches to evaluation and impact (Befani, et al., 2015) it is argued that understanding a **plurality of perspectives beyond those of the implementers and beneficiaries** is essential for rigorous evaluations based on concepts of ‘systems’. In this issue, Williams (2015) states, ‘Much of the discourse around impact evaluation is the contribution of an intervention to so-called intended impacts. A perspectives orientation begs the question: ‘Whose intended impacts are being used from what framing?’ (12).

The frameworks for guiding and assessing market systems interventions are almost always decided primarily by programme staff and donors. The perspectives of both beneficiaries and non-beneficiaries are also needed to understand the value and relevance of initiatives. After all, ‘they are the only people who have expert knowledge of the complexities they experience’ (Chambers, 2015). This needs to include (a) the market actors who are expected to change their behaviours, (b) the people living in poverty who are expected to be affected by these changes, and (c) the people living in poverty who are not necessarily expected to benefit from changes but may still be affected. However, people in groups (b) and especially people in group (c) are rarely involved in debating intervention strategies or analysing assessment findings. In fact, when people do not participate in an intervention, it is sometimes assumed that it is because they are ignorant of what is good for them, or because they ‘lack the right mindset’ (individual working for an M4P donor). These assumptions arguably privilege the perspectives of the programme implementers and donors over the perspectives of potential beneficiaries such as non-participating farmers.

In evaluations, the ability of people living in poverty to conduct their own analysis is often overlooked: ‘we go to the farmers and look at the productivity, and then we kind of calculate the profitability for the farmers to get an estimate of what it was and compare it to the previous season.’ Note here that farmers are not asked what has changed for them, but rather profits are calculated on their behalf. However, studies have shown that participatory approaches are able to provide not just qualitative data and analysis, but also quantitative data and analysis, such as through participatory statistics (Barahona and Levy, 2005;

Holland, 2013). Further, the process of analysis, or making sense of data, is arguably just as important for validity as the data itself (Pawson, 2003), and this process depends largely on the perspectives and values of the people involved.

The practice of incorporating the perspectives of beneficiaries (e.g. farmers included in a contract farming scheme) and non-beneficiaries (e.g. farmers which have not been included, farmers which have opted out, non-farming households, local traders, household members of included farmers, etc.) in assessment work is not without its challenges (Burns and Worsley, 2015). Mitigating expectations about what the programme can achieve and respecting commercial confidentiality are two specific challenges that market systems development programmes may experience in attempting to incorporate a broader range of perspectives in their assessments. However, these challenges can also occur with other assessment approaches, from surveys to interviews, and the potential increase in quality of evidence generated by widening out the perspectives included in programme assessments is arguably worth efforts to overcome and mitigate these challenges.

One online discussant noted that assessing market development initiatives ‘require[s] a highly political and subjective process of agreement *between practitioners and donors*. Studies, science, statistical analysis, historical analysis, typologies and all things rational must prevail, but at some point, we must just surrender to the belief that certain types of changes in networks and cognitive structures are a “good” thing and will take a certain type of market system into a future of higher inclusion, adaptability, productivity, efficiency, etc.’ [emphasis added]. It is this author’s opinion that in order to improve the quality of evidence about market systems approaches, the ‘agreement’ needs to also be made with those people living in poverty who will be affected by the intervention, including but not limited to the intended beneficiaries.

3. Suggestions for feasibly generating rigorous evidence for market systems approaches

Building on the discussion above, this section provides a set of recommendations about how to improve the rigour of evidence in market systems initiatives. It offers a set of principles which could be used to guide monitoring and evaluation efforts. This is followed by a discussion of how to balance rigour with feasibility.

a. Four principles for improving the rigour of assessments

Rather than advocate for a particular methodology, which could serve to replicate the problems already experienced in the debates about methods and methodologies, this section provides a list of overarching principles or requirements for rigorous market systems assessments. Based on the consultations and literature analysed here, the following four elements are proposed for assessment efforts¹³ that can produce rigorous and credible evidence of the effects of market systems interventions:

¹³ This can include monitoring and evaluation or impact evaluation efforts, depending on the scope of the programme. Due to the scope of this paper, the separations between monitoring, evaluation, and impact evaluation are not discussed.

- (i) **builds on and incorporates the pluralistic perspectives, experiences, and values of the beneficiary and non-beneficiary people** living in the areas of an intervention (i.e. being participatory), with a particular focus on those living in poverty
- (ii) **captures the complex system changes of people living in poverty.** Again, it is worth noting that while some elements of complex systems analysis are already taking place in market systems initiatives, the focus of such analysis is on the level of the market. What is proposed here is a stronger focus on the systems dynamics of people living in poverty: that is, how the effects of the intervention (i.e. market changes) affect and are affected by the realities of people living in poverty. In addition to bringing assessments more in line with concepts of complex systems, this approach can potentially mitigate the issue of time lag in that it can provide an indication of what is shifting in the systems of people living in poverty, possibly before a (longitudinally) measurable difference emerges.
- (iii) **inquires into the specific areas which the programme team or evaluators anticipate will be of interest** (i.e. testing the programme theories of change and providing information about achievements and results), and crucially,
- (iv) **captures changes beyond what the programme team or evaluators may have anticipated** (i.e. it captures the broader ‘emergent’ changes associated with an intervention that may not have been anticipated, such as changes to local food prices, land access, gender roles, soil quality, rural-urban migration patterns, etc.. These realities are dynamic and multidimensional and changes cannot necessarily be predicted by programme teams; therefore, it is possible that some of the most significant changes cannot be captured using indicators related to goals or risks. By incorporating a wider range of perspectives in point (i) and focusing on the wider systems of people living in poverty in point (ii), assessments will be better able to surface these changes which may not have been on their radar.

By approaching assessments in a way that captures a plurality of perspectives, captures the complex changes at the level of the poor and captures unanticipated issues as they emerge, assessments could achieve a breadth and depth that does not come from studies which focus on a narrow evaluation question such as, ‘has production increased’. They could allow for unanticipated changes and dynamics to surface from people’s own experiences and observations could help to address the issue of generating a greater understanding of a programme’s effects, including how the programme affects the poorest and most marginalized segments of the population, even (or particularly) when these people are not targeted (either directly or indirectly). This fits well with market systems initiatives which often affect change in an indirect way.

Examples of methodologies which tend to incorporate these principles include Participatory Systemic Inquiry, Systemic Action Research (Burns and Worsley, 2015; Burns, 2007), and Most Significant Change (Davies and Dart, 2005). There are undoubtedly other methodologies and approaches that can also be applied to address these four issues. Of course, the way in which an approach is implemented is crucial: methods with great potential can be implemented very poorly. It is important to note that this piece is not advocating for a particular methodology, which could serve to replicate the problems already experienced in the debates about methods and methodologies. Rather, these approaches are offered as

examples which have the potential to address the limitations in existing evidence for market systems initiatives based on the analysis offered here.

It is also important to note that assessments which align with these principles do not necessarily need to incorporate new methods; it could simply be a matter of applying an existing method in a new way, such as by incorporating a wider set of stakeholders and restructuring interviews and group discussions so as to encourage the emergence of other issues which are occurring in the lives of people living in poverty.

b. Putting it into practice: finding balance between rigour and feasibility

For good reason, practitioners, evaluators and donors are wary of having to measure absolutely everything. At the end of the day, measurement needs to fit into limited programme resources, which includes time and financial considerations. ‘What you have to avoid is unmanageable research so that people start doing it’ (M4P Team Leader Interviewee). So the question is how to practically generate rigorous evidence without the process being too cumbersome or resource intensive.

Again, one unique characteristic of market systems initiatives is that they typically seek to assess changes in (complex) market systems, on top of changes at the beneficiary level. With these assessments keeping programme teams busy, there is understandably limited capacity to add further areas of focus. Nonetheless, many practitioners also aim to look out for unintended effects, and the DCED Standard also encourages, but does not require, this (Recommendation 3.4i in control point on measuring changes in indicators).

When considering what to assess beyond targets, practitioners pragmatically try to identify what could be important to measure. One online discussant involved with the DCED Standard noted that ‘by encouraging people to think deeply about their theory, we enable them to look for unexpected effects.’ In an interview, a practitioner stated, ‘If you know there is an environmental hazard you should build it into the MRM, you should look at gender because of household activities, but you should not put everything on the list. You need to be informed about risks and check those risks.’

While it is important to try to anticipate and track changes other than those which are ‘intended’ as outcomes and impacts through the use of theory and relevant experience, it is also essential that practitioners are open to the fact that they might not be able to anticipate all of the potential effects. This element of unpredictability is one of the main characteristics of complex systems (Heylighen, 2008).

This topic was raised during the online discussion, but there was concern about what was practical given the existing demands on practitioners. One participant stated that, despite the fact that capturing unintended effects is encouraged in the DCED Standard Guidelines, given the demands of assessing systemic change in markets, ‘there are few if any major examples’ [of programmes assessing complex levels of changes]. Another participant noted that one market systems programme was taking an interesting approach to assessing ‘graduation’ from poverty, but it was unclear whether this greater focus on poverty came at the expense of assessing systemic changes in markets.

Discussants asked whether it would be feasible to assess both systemic changes in markets as well as the ‘social, economic, cultural and environmental systems of the poor.’ ‘Can we

look at both at the same time?’ One participant asked whether something in the existing assessment efforts could be ‘replaced to make room’ for assessing multidimensional changes beyond what is targeted, particularly at the level of people living in poverty. Alternatively, it was suggested to advocate for more resources for project monitoring, or to shift the assessment of multidimensional effects to academics, with their own budgets rather than including this in the tasks for internal programme monitoring.

Approaches which incorporate of a wider set of perspectives in assessments—and which privilege the perspectives of people living in poverty—can help enable programme teams to better understand which effects are significant which are marginal. This can help programme teams capture the wider effects of their programmes—and therefore increase the rigour of their assessments—while at the same time eliminating the pressure to include ‘everything but the kitchen sink’ in their efforts.

In addition, approaches which align with the principles presented in the previous section such as Participatory Systemic Inquiry and Most Significant Change can be implemented in a way which requires relatively few resources and time when compared to other approaches such as surveys and quasi-experimental studies¹⁴. They often entail gathering people together in workshops, which represents a more efficient way to engage with people than individual engagement. Participatory approaches often engage community members in data collection and analysis itself, which could also reduce the amount of input required by project monitoring and evaluation teams. Developing in-house capacity for these types of approaches could also mean less reliance on costly external consultants, and, if an appropriate learning culture exists, can help facilitate improved responsiveness to social, economic and environmental realities of people living in poverty.

4. Concluding remarks

Calls for rigour are often misinterpreted as calls for certain methods which are deemed as rigorous; however, what this piece argues, drawing on both literature and the perspectives of people working in the market systems field, is that the concept of rigour goes beyond specific methods to include what questions are asked and not asked, who is involved in framing and participating in assessments, the use of appropriate methodologies to answer these questions, and the existence of learning cultures that facilitate generating and responding to evidence.

Asking the right **assessment questions** is crucial to generating rigorous evidence about the effects of an intervention. It matters little if a programme is able to confidently assess a given change if other significant effects of the programme, and other framings of those effects, are not also included. This piece argues that studies need to go beyond the intended results to

¹⁴ A full analysis has not been completed to compare the cost-effectiveness of approaches which fulfil the four principles outlined in the previous section, and to this author’s knowledge there are no overall studies which compare the cost effectiveness of participatory assessments to non-participatory assessments. However, these issues are discussed by Guijt 2014 (http://devinfoolive.info/impact_evaluation/ie/img/downloads/Participatory_Approaches_ENG.pdf) and FAO 2003 (<ftp://ftp.fao.org/docrep/fao/007/ad817e/ad817e00.pdf>)

assess the wider multidimensional effects of the programme, particularly in relation to people living in poverty. To date, the focus on systemic effects has largely been at the level of markets, while assessments of effects experienced by people living in poverty are often linearized, focusing primarily on production, income or access to goods. Assessments at the level of people living in poverty are also typically limited to intended beneficiaries, who are often not the poorest, and assessments generate little understanding about the effects that other people living in poverty (i.e. non-beneficiaries) may have experienced. Thus, it is recommended that assessments of market systems initiatives attempt to assess the complex social, economic and environmental systems of people living in (extreme) poverty, including beneficiaries and non-beneficiaries.

In terms of **assessment approaches** which can achieve this, approaches which incorporate a diversity of perspectives—including but not limited to **beneficiaries and non-beneficiaries living in poverty**—can provide an understanding of the different ways in which a programme (or intervention) might be playing into economic, social, environmental, health, food security, and other systems of people living in poverty. By broadening out the scope of assessments to consider the systemic effects experienced across different dimensions and by different groups of people, and making explicit the different values and stakes in a given intervention, programmes can rigorously test the validity of their programme logic and theories of change. As Williams states, ‘a multiframing approach can help explain much more clearly why a programme succeeded and how to improve it than the single framing based on the ‘intended impact’ (Williams, 2015:12).

Further, **facilitating a learning environment and culture** is crucial. Without it, rigorous evidence will not be generated or shared in order to inform decision making. Practitioners indicated that there is a need for monitoring and evaluation to be championed by managers: ‘There is always something else more urgent, no one likes it, it is a scary thing and so people turn the other way, there is always another report that needs to happen first’. Equally, donors play a major role in cultivating an enabling environment for the production and use of rigorous evidence. The ability to generate and act upon rigorous evidence partly depends on how receptive donors are to changes in programme design and direction based on new understandings of a programme’s relevance and effectiveness; and how well they communicate this receptivity to managers.

References:

Argyris, C. & Schon, D. (1974). *Theory in Practice*. San Francisco: Jossey-Bass

Atkins, D., Eccles, M. et al., (2004) Systems for grading the quality of evidence and the strength of recommendations 1 : critical appraisal of existing approaches. *The Grade Working Group*. ‘BMC Health Services Research.’ 4(1): 38.

Bagshaw, S. and Bellomo, R. (2008) The need to reform our assessment of evidence from clinical trials: A commentary. ‘*Philosophy, Ethics, and Humanities in Medicine*.’ 3:23 (doi: 10.1186/1747-5341-3-23)

Barahona, C. and Levy, S. (2005) ‘The best of both worlds: producing national statistics using participatory methods’ Q-squared Working Paper No. 11, Toronto: Centre for International Studies

- Befani, B., Ramalingam, B. And Stern, E. (Eds.) (2015) *Towards Systemic Approaches to Evaluation and Impact* IDS Bulletin 46:1
- Burns, D. and Worsley, S. (2015) *navigating Complexity in International Development: Facilitating sustainable change at scale*, Rugby: Practical Action
- Burns, D. (2007) *Systemic Action Research: A strategy for whole system change*, Bristol: Policy Press
- Chambers, R. (2015) 'Inclusive rigour for complexity' *Journal of Development Effectiveness*. <http://dx.doi.org/10.1080/19439342.2015.1068356>
- Cook, T.D. and Campbell, D.T. (1979) *Quasi-Experimentation: Design and Analysis for Field Settings*. Rand McNally, Chicago, Illinois
- Dalziel, M. et al., (2012) Impact of government investments in research and innovation: review of academic investigations. 'Optimum Online.' 42/2, June 2012
- Davies, R. And Dart, J. (2005) 'The 'Most Significant Change' (MSC) Technique: A guide to its use' Available from <http://www.mande.co.uk/docs/MSCGuide.pdf>
- DCED (Donor Committee for Enterprise Development) (2015) DCED Standard for measuring results in private sector development: Control points and compliance criteria. Version VII, April 2015. Available from <http://www.enterprise-development.org/page/download?id=1842> (Accessed October 2015)
- Du Toit, A. (2012) 'Making sense of 'evidence' Notes on the discursive politics of research and pro-poor policy making' Institute for Poverty, Land and Agrarian Studies (PLAAS) *Working Paper 21*, University of the Western Cape: Bellville
- Fold, N., (2009). Finding zones of convergence in a world of continental drift. *Singapore Journal of Tropical Geography*, 30(1), pp.13–17.
- Garcia, J. And Zazueta, Z. (2015) 'Going beyond mixed methods to mixed approaches: a systems perspective for asking the right questions.' In Befani, et al. (Eds.) *Towards Systemic Approaches to Evaluation and Impact* IDS Bulletin 46:1
- Gilligan, C. (2012). 'Immigration controls and the erosion of popular sovereignty.' GRITIM Working Paper Series Number 10, Winter 2012.
- Heylighen, F. (2008) 'Complexity and Self-organisation' in Bates, M. and Maack, M. N. (Eds.) *Encyclopedia of Library and Information Sciences*. Taylor & Francis
- Holland, J. (2013) *Who counts? The power of participatory statistics* Rugby: Practical Action Publishing
- Howlett, P. and Morgan, M.S. (Eds.) (2010) 'The nature of evidence: how well do 'facts' travel?' Final report 2004-2009. London: LSE.
- Hummelbrunner, R. (2015) 'Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence' in Befani, et al (Eds.) *Towards Systemic Approaches to Evaluation and Impact* IDS Bulletin 46:1

- Jones, H. (2011) 'A guide to monitoring and evaluating policy influence' Overseas Development Institute Background Note
- Imam, I., LaGoy, A., and Williams, B. (2006) 'Introduction' in Williams, B. and Imam, I. (Eds.) 'Systems concepts in evaluation: an expert anthology.' American Evaluation Association.
- Kessler, A. (2014) 'Assessing Systemic Change: Implementation guidelines for the DCED Standard' Available from <http://enterprise-development.org/page/download?id=2113>
- Kuhn TS (1962) *The Structure of Scientific Revolutions*. University of Chicago Press: Chicago, USA
- Levitt, R. (2013) 'The challenges of evidence' *Provocation paper for the Alliance for Useful Evidence, NESTA*
- Morse, JM, and Richards, L. (2002) *Readme First for a User's Guide to Qualitative Methods*. Thousand Oaks, CA: Sage Publications.
- Nippard, D., Hitchins, R. And Elliot, D. (2014) 'Adopt-adapt-expand-respond: a framework for managing and measuring systemic change processes' The Springfield Centre Briefing Paper March 2014
- Nutley, S., Powell, A. and Davies, H. (2013) 'What counts as good evidence? Provocation paper for the Alliance for Useful Evidence' Research Unit for Research Utilisation (RURU) University of St Andrews, Alliance for Useful Evidence
- OECD-DAC (2010) 'Quality standards for development evaluation' *DAC Guidelines and Reference Series*
- Oya, C., (2012) Contract Farming in Sub-Saharan Africa: A Survey of Approaches, Debates and Issues. *Journal of Agrarian Change*, 12(1), pp.1–33.
- Patton, M. Q. (2015) *Qualitative Research & Evaluation, Integrating Theory and Practice* Fourth Edition, SAGE. Excerpt available from http://betterevaluation.org/blog/anecdote_as_epithet (Accessed September 2014)
- Pawson. R, (2003) 'Assessing the quality of evidence in evidence-based policy: when, why, how and when.' ESRC Research Methods Programme Working Paper No 1
- Pawson R., Boaz A., Grayson L., Long A. & Barnes C. (2003) 'Types and Quality of Knowledge in Social Care.' Social Care Institute for Excellence, London
- Pawson, R. (2007) *Causality for Beginners*. NCRM Research Methods Festival 2008, Leeds University. Available from: <http://eprints.ncrm.ac.uk/245/>
- Perrin, B. (2013) 'Overall evaluation of the M4P delivery model: proposed approach' *Service Request Report*, Prepared for DFID, International Organisation Development Ltd
- Ruffer, T. And Wach, E. (2013) 'Review of Making Markets Work for the Poor (M4P) Evaluation Methods and Approaches' DFID Working Paper 41
- Sackett D and Wennberg, J. (1997) Choosing the best research design for each question: It's time to stop squabbling over the 'best' methods. *British Medical Journal*: 315:1636

Sarwar, R. (2015) 'Bangladesh Health Service Delivery: Innovative NGO and Private Sector Partnerships' in Thorpe, J. and Wach, E. (Eds.) *Business, State and Society: Changing Perspectives, Roles and Approaches*, IDS Bulletin 46:3

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

Stern, E. Stame, N. Mayne, J. Forss, K, Davies, R. and Befani, B. (2012) 'Broadening the Range of Designs and methods for Impact Evaluations. Report of a study commissioned by DFID.' DFID Working Paper 38.

Vicol, M. (2015) 'Corporatisation of rural spaces: contract farming as local scale land grabs in Maharashtra, India' *Land grabbing, conflict and agrarian-environmental transformations: perspectives from East and Southeast Asia*, Conference Paper No. 28, Chiang Mai

Wach, E. (2011) 'Measuring the Inclusivity of Inclusive Business' *IDS Practice Paper*, No. 9. Brighton: IDS. Available from: <http://opendocs.ids.ac.uk/opendocs/handle/123456789/3149>

Williams, B. (2015) 'Prosaic or profound? The adoption of systems ideas by impact evaluation' in Befani, et al (Eds.) *Towards Systemic Approaches to Evaluation and Impact* IDS Bulletin 46:1