

Improving the Evaluability of INGO Empowerment and Accountability Programmes

Abstract This CDI Practice Paper is based on an analysis of international NGO (INGO) evaluation practice in empowerment and accountability (E&A) programmes commissioned by CARE UK, Christian Aid, Plan UK and World Vision UK. It reviews evaluation debates and their implications for INGOs. The authors argue that if INGOs are to successfully ‘measure’ or assess outcomes and impacts of E&A programmes, they need to shift attention from methods to developing more holistic and complexity-informed evaluation strategies during programme design. Final evaluations or impact assessments are no longer discrete activities, but part of longer-term learning processes. Given the weak evaluation capacity within the international development sector, this CDI Practice Paper concludes that institutional donors must have realistic expectations and support INGOs to develop their evaluation capacity in keeping with cost–benefit considerations. Donors might also need to reconsider the merits of trying to evaluate the ‘impact’ of ‘demand-side’ NGO governance programmes independently of potentially complementary ‘supply-side’ governance initiatives.

1 Introduction

If international NGOs (INGOs) are to successfully ‘measure’ or assess outcomes and impacts of empowerment and accountability (E&A) programmes, they need to shift their attention from indicators and data collection tools to a more holistic approach to thinking about appropriate monitoring and evaluation (M&E) strategies and systems. If they are to lead or commission evaluations and impact assessments that both generate evidence of the desired quality, and are consistent with their organisations’ participatory values, they need to start developing monitoring, evaluation and learning (MEL) strategies at the programme planning and budgeting stage.

This CDI Practice Paper is based on a scoping and analysis of INGO evaluation practice commissioned by CARE UK, Christian Aid, Plan UK and World Vision UK using funding from their DFID Programme Partnership Agreements (PPAs). We reviewed a range of E&A evaluation documents relating to 16 programmes and projects, varying from a randomised control trial designed and

implemented by academics, through multi-country INGO governance portfolios, to modest country office pilots of participatory M&E for internal learning.¹ In doing so, we also synthesised and brought to bear lessons from recent MEL literature on the INGOs’ practice.

The final product was not, as originally envisaged by the INGOs, a selection of ‘best practice’ tools or indicators, but a paper on improving the evaluability of INGO E&A programmes (Shutt and McGee 2012). From this, we have derived two CDI Practice Papers:

- the present paper reviews, in turn, the state of MEL debates, their implications for INGOs and emerging practice, a summary of guidelines and a brief conclusion; and
- CDI Practice Paper 01 Annex sets out in full detail the guidelines summarised briefly in this paper.

A word on terminology. Throughout the two CDI Practice Papers, rather than using the terms ‘measurement’ or

'impact evaluation', we have tended to use 'assessment' and 'exploration'. This is partly to avoid the strictures of the narrowest definition of 'impact evaluation', to which a small proportion of donors and evaluators subscribe, where the most 'robust' methodologies are used to explore causality as a basis for evidence for influencing policy and spending decisions. And it is partly because our view of the field suggests that sometimes 'assessment' and 'exploration' are more accurate and appropriate ways than 'measurement' to describe what is being done, and indeed what can be done, in the context of E&A programmes.

2 The state of MEL debates

While international development NGOs and their southern partners are eager to demonstrate their effectiveness, in the 2010s they are struggling to respond to ever more stringent requirements from their institutional donors to demonstrate the impact of their activities. These requirements arise from the exigencies of the results agenda and evidence-based turn in aid policy, in the UK, USA and other OECD donor countries. Development INGOs encounter these challenges in designing and implementing their internal M&E systems, and also in commissioning external, independent evaluations of their programmes funded by institutional donors.

Over decades, INGOs' impact claims have tended to err on the side of vagueness, and have blurred the fact that the makers of these claims were highly positioned actors acting in the interest of their own organisations, constituencies, sector and marginalised target groups, as well as, or more than, in the interests of methodological rigour. INGO staff are now keenly aware of the need to achieve rigour in assessments of their own impact, and of the risks entailed in impact evaluation that smacks of hubris or self-justification.

Thematically speaking, the field where INGOs – and their official aid and philanthropic donor agencies too – are having most trouble is in the field of governance programming. This is hardly surprising. The governance field is broadly regarded as a 'complex' one for programming, frequently contrasted with fields such as malaria prevention or vaccination where – providing donors are supporting direct delivery models – results can be counted and attributed as unambiguously as beans. (Whether the bean count gives a full account of change being achieved by programmes that aim to develop the long-term capacity of health systems to deliver vaccinations, which usually also involve messy governance components, is quite another question.)

A study was commissioned by DFID's Evaluation Department in 2011 with the specific mandate of 'broadening the range of designs and methods for impact evaluation'. It was born of a rising concern that while

experimental impact assessment methods imported from the medical sciences were catching on and spreading across the development and aid fields, DFID and its partners, peers and grantees were lacking evaluation methods that fitted comfortably with a growing proportion of their programmes. Particularly noted was the lack of suitable approaches for assessing the impact of governance programming in such sub-fields as 'strengthening democracy and accountability; accountable and responsive government; security and conflict prevention; combating gender-based violence; citizen empowerment and community action'.²

The DFID study coincided in time with the first large-scale review of the effectiveness and impact of accountability and transparency initiatives, commissioned by the T/A Initiative.³ Also around the same time, several researchers and practitioners in the development and aid fields published critical, thought-provoking, questing work about the challenges of assessing impact in governance programmes or what many of them called 'complex' programmes, in which they include governance programmes (Ramalingam *et al.* 2008; Hughes and Hutchings 2011; Roche and Kelly 2012; Barder and Ramalingam 2012).

These sources show a remarkable degree of consensus on the nature of the 'problem' at hand and largely concur on a set of key points which are highly relevant to this CDI Practice Paper.

Firstly, they agree, we are in the grip of an 'Evidence-Based Policy Movement' (a phrase used by Stern *et al.*). This has thrown into sharp relief the generally weak state of evaluation and impact assessment within the development sector. Particular weaknesses identified relate to evaluation designs and the quality of analysis, which have tended to overlook the importance of context and power relations that are of particular concern in E&A interventions.

Acknowledgement of the weak state of current practice and concerns about improving it have moved evaluation debates beyond discussions of differences between quantitative and qualitative or participatory methods. Arguably, in the past a focus on methods has diverted attention from deeper and more significant differences among the evaluation community. These are better described in terms of 'causal wars' and are both political and methodological in nature (Roche and Kelly 2012). Those determined to attribute impacts to their interventions have tended to favour experimental and quasi-experimental evaluation designs. This is because such designs are believed to be superior when it comes to demonstrating that development initiatives 'cause' impacts within short-term timeframes and thus tend to fit with political demands for quick evidence of results. On the other hand, those who see development agencies playing a relatively small role in complex, long-term social change processes have been driven by different imperatives. They tend to be less

Box 1 Difference between methodological 'design' and 'methods'

Design is defined as the overarching logic for evaluations that includes: evaluation questions, theory used to analyse data, data and use of data.

Methods are defined as approaches to data collection and measurement tools and statistical analysis. (Adapted from Stern *et al.* 2012)

concerned about establishing causal relationships and demonstrating that a certain outcome can be attributed to a particular intervention, and more interested in using approaches that are contextually embedded and that encourage learning by different stakeholders. All too often these latter approaches have been insufficiently critical and confounded by weak methodological design.⁴

The recent contributions to the debate, particularly Stern *et al.* (2012), comment on an increased willingness by practitioners who have previously held quite different positions to accept the need for more pluralist designs and methods. They distinguish usefully between evaluation designs and methods; see Box 1.

The same methods can of course be deployed within different designs. However, choices between different methods are driven by design, because it is the objective and questions of the research or evaluation that influence researchers' and evaluators' ability to assess 'causal

inference' and attribution or contribution, rather than the methods. Underpinning different designs and design decisions are different epistemologies and methodological paradigms, which are based on different premises or 'belief systems' about the nature of knowledge and truth claims. While a detailed discussion of these lies beyond the scope of this CDI Practice Paper, it is important to note that such deep, often buried, underlying 'belief systems' have profound implications for practice.

Tensions between 'learning' and 'donor accountability' evaluation objectives are well established in the literature.⁶ A-theoretical experimental and quasi-experimental designs that aim to answer the question 'Did the programme work?' without exploring how, why, and in what contexts are often better suited to evaluations trying to demonstrate accountability to donors, particularly donors who are spending tax-payers' money, rather than fostering learning. In practice, the distinctions are somewhat more complex. In an evidence-based policy paradigm, experimental methods are often used judiciously for the impact evaluation of a small proportion of donor programmes, to inform learning about whether policy approaches 'work' or not. The increasing use of hybrid designs that integrate programme designs based on theories of change, with experimental evaluation designs, means that many offer more learning possibilities about how and why impact does or does not result from interventions than hitherto assumed.

New definitions and understandings of 'gold standards' in terms of rigour in evaluating development initiatives are

Table 1 Evaluation designs and basis for causality claims⁵

Evaluation design approaches	Basis for claims of causality
Experimental Randomised Controlled Trials (RCTs), quasi-experiments, natural experiments	Counterfactuals; the co-presence of cause and effect
Statistical including descriptive statistical modelling and longitudinal studies	Correlation between cause and effect or between variables
Theory based Using a theory of change (ToC), process tracing, contribution analysis and impact pathways	Identification or confirmation of causal processes or chains and supporting factors and mechanisms at work in the context. May include some statistical analysis
Case based Interpretive – naturalistic, grounded theory or ethnography	Comparison across and within combinations of causal factors
Participatory Participatory evaluation for empowerment.	Validation by participants that their actions and experienced effects are caused by the programme.
Agency-led Learning by doing, collaborative action research	Involves adoption, customisation and commitment to a goal
Synthesis Meta analysis, narrative or realistic-based synthesis (Pawson 2002)	Accumulation and aggregation within a number of perspectives above

(Adapted from Stern *et al.* 2012)

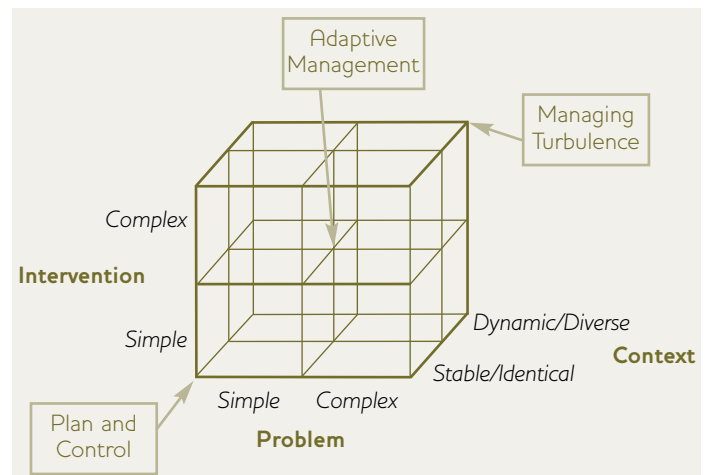
coming to light as evaluators recognise the importance of evaluation designs and the potential of non-experimental designs to explore causal mechanisms. What defines a 'gold standard for impact assessment' is no longer the scope it offers for establishing causal relationships, but rather how appropriate the methodological design is given the emphases in the questions asked (do they emphasise questions of *whether*, *how much*, or *how* cause led to effect?), and the attributes of the programme in question. There is now greater awareness of trade-offs between the scope and complexity of programmes on the one hand, and the confidence with which causality can be established, on the other. If these lessons are to be usefully applied to practice, challenges need anticipating and mitigating at programme design stage, not at evaluation stage. Since so many of the lessons relate to programme design, by the ex-post evaluation stage it is too late to incorporate many of them.

Several of the broadly like-minded sources we cite above note the tensions between complex programme environments and demands for neatly demonstrated results. Barder and Ramalingam (2012) go a useful step further than lamenting these tensions. They argue that there is no contradiction between recognising complexity in the programme environment, and focusing on and demonstrating results. In fact, they hold, 'Complexity provides a powerful reason for pursuing the results agenda, but it has to be done in ways which reflect the context'. They present Ramalingam's 'complexity-aware approach to results', based on the proposition that we need to get a better handle on the *nature* of the programme in question, the *interventions* being implemented (including the complexity of relationships within them), and the *context* of the intervention – which affects not only what change is possible, but also the feasibility of evaluation.

Barder and Ramalingam go on: 'This gives us three dimensions – ranging from simple problems and interventions in stable contexts through to *complex interventions in diverse and dynamic contexts*' (italics added). The italicised phrase closely fits E&A programmes which, by dint of their objectives, are often developed for and unfold in a range of governance contexts characterised by polarised power relations and fluid and unpredictable sociopolitical dynamics. They continue:

'Down in the bottom left-hand corner are simple problems and stable settings. This is where *'Plan and Control'* makes most sense. Traditional results-based management approach, the more conventional unit-cost based value [sic] for money analyses and randomised control trials work especially well... At the top right we have complex problems, complex interventions in diverse and dynamic settings. (A lot of donor work in fragile states and post conflict societies are in this corner.) Here the goal is *'Managing*

Turbulence'. In this space, everything is so unpredictable and fluid that planning, action and assessment are effectively fused together... In between is what we have called the zone of *'Adaptive Management'*. Here we may find ourselves managing a variety of combinations of our three axes. In our view, the vast majority of development interventions sit in this middle ground' (Barder and Ramalingam 2012).



Contributions like Barder and Ramalingam's make it legitimate and appropriate to recognise that many aspects of development and humanitarian programme contexts are beyond the control of an INGO programme or, indeed, any other kind of purposeful intervention; and that complex problems which require complex interventions, such as attempts to redress power imbalances and get historically unaccountable governments to become more accountable, compound this difficulty. This helps guide choices of evaluation approach according to a common-sense analysis of what we can and cannot purposefully manage or measure.

3 Implications for INGOs and emerging practice

Bearing in mind the evidence of current practice that we analysed during our scoping, what do these general MEL debates imply for INGOs implementing E&A programmes?

Resist 'one-size-fits-all' approaches to E&A MEL In a world which, many development practitioners agree, is messy and non-linear, it is important to avoid both reductionist, 'one-size-fits-all' approaches to assessing results and excessive pessimism about possibilities for learning and adaptation. This points us towards a more adaptive approach, and to experimenting with hybrid designs (including elements of thinking from theory-based approaches, participatory and case study designs) to monitor emergent progress and adapt to a changing context. Real-time learning is essential to check the relative effectiveness of different approaches and scaling down or stopping those that are ineffective (Barder and Ramalingam 2012).

Think about how social change actually happens, with or without external interventions Barder and Ramalingam's various dimensions resonate with Reeler's (2007) models of social change, derived from extensive practice and systematic field observation: emergent change, transformative change and projectable change. 'Projectable change' leaps out at development practitioners as what we attempt to do, often unrealistically. 'Emergent change' chimes with the emergent and uncontrollable qualities of the realities we work in. 'Transformative change' strikes some of us as the most far-reaching effective type of change we have ever observed or experienced.

Logically, how one goes about trying to engage with social and political change processes to make them happen more, less, or differently, should be informed by how one understands social change as actually happening. External empowerment and accountability interventions need to be viewed as coming into a situation in flux, in which any number of change processes, of any mix of Reeler's three types, are ongoing, interwoven with each other. If the external actor seeks 'socially progressive' outcomes, it will want to engage with this fluid field of change dynamics and help it onwards in a certain direction. It might do that in a range of ways: by blocking one particular change process, starting up an entirely new change process, engaging with a change process tangentially so as to speed it on its way.

By implication, to be maximally effective – particularly in terms of benefiting vulnerable people – the external agent has to have some more or less well-informed notion of what change dynamics are going on and what drives and shapes them, in this place, at this time. S/he has to design the intervention according to that notion and be prepared to modify it as the local contemporary change dynamics unfold – which, if left to their own devices, are probably more likely to happen in an 'emergent' or 'transformative' way than in a 'projectable' way.

More attention to context and power analysis The problem of many development and aid programmes since aid began is that they are premised on projectable models of change, attributing to the programme environment qualities and dynamics that it does not have. Various remedies for this mismatch between aid programming and real life have been tried since the early 1990s when a series of development studies 'classics'⁷ made it increasingly untenable to ignore the mismatch. Transposed into the field of empowerment and accountability programming, the mismatch translates into 'power blindness'. This blindness consists of a failure to recognise that what is at stake is the inequitable power relations between actors, and that both 'empowerment' and 'accountability' are fundamentally about shifting power relations in ways that may benefit the programme beneficiaries but may also disempower those who have

failed to account to them and have benefited from their powerlessness. Power needs to sit squarely within our thinking around E&A programmes and the ways we try to assess their impact must therefore reflect political as well as technical considerations.

Shift from assuming attribution to considering contribution Projectable models of social change processes have been overplayed especially by aid agencies needing to show that they can make major contributions to social change in short periods of time. These actors need to get better at asking what else besides their project may have facilitated or hindered change: local political dynamics? religious fervour? demographic dynamics? Most development INGO M&E strategies and systems, including those reviewed as part of this scoping, have started from the 'projectable change' ('Working with a plan') understanding of truth, evidence, knowledge and reality, rather than grounded contextual understandings which are closer to the 'emergent' understanding ('We make our path by walking it') (Reeler 2007). They need to shift their focus from seeking attribution ('What did our project achieve?') to considering contribution ('Has anything changed and did we have anything to do with it?').

The above will require a shift in emphasis from focusing on indicators, data collection tools and methods, to more critical thinking about choosing MEL strategies and designs that enable appropriate exploration of cause and effect relationships. Those involved in evaluation and impact assessment of E&A initiatives not only face the challenge of identifying and assessing meaningful measures of change. They also need to reflect more on whether INGO interventions are necessary and/or sufficient – in the absence of other factors – to explain any social, political and economic empowerment of different groups of people observed. They will need to ask whether, and in what contexts and circumstances, different types of citizen empowerment leads to meaningful changes in the responsiveness and institutionalised accountability behaviours of state actors at different levels, private sector service providers, and indeed themselves. Developing approaches and systems to assess these changes is vital if impact evaluations are to detect and explain the longer-term effects of programmes on people's wellbeing.

In assessing contribution, there is no expectation of accurate quantification. Instead, shifting the focus to contribution can be seen as an opportunity for those working at the frontline to apply their contextual knowledge to understanding change or lack of it, and improving programmes in an evolving way that has been discouraged – if not 'disallowed' – in traditional 'projectable' results-based approaches to M&E.

Start evaluation planning earlier Like Stern *et al.* (2012), we found a lack of baseline and process monitoring data

has limited what final evaluations, conducted within time constraints, can reliably conclude or achieve. This suggests an urgent need to start developing MEL strategies earlier in the programme's life – at the design stage.

Less data and more real-time analysis Recent comments by DFID evaluation specialists, Nick York and Caroline Hoy, on Oxfam GB's attempts to measure its effectiveness are telling.⁸ They caution against investing time and effort in collecting and analysing lots of quantitative data using counterfactuals in final evaluations when other approaches to understanding change may be more appropriate.

'One key (Oxfam GB) assumption is that by doing more work and collecting more data... from comparison sites... they will be able to understand and demonstrate impact. Actually, based on discussions we have had... recently in DFID, we have started to ask a different sort of question. In some types of programmes, more data and more work may not be the solution – more innovative methods and approaches to understanding impact can be required and if the programme itself develops as you implement it then the goal posts are continually shifting too.'⁹

Improve the quality of analysis and learning Like the evaluations assessed by Stern *et al.* (2012) and McGee and Gaventa (2011), INGO evaluations have tended to be better at descriptive statistics and generating general lessons than analysing causal mechanisms and undertaking the type of learning directed by a 'theory of change'-based interrogation. This is not only because of a lack of monitoring data. Even when such data exists, it is seldom used for triangulation and interrogation of how and why results happened. Instead of asking 'Are we doing things right?', we need to ask 'What do our findings suggest about our assumptions and theories about how change happens?'

Long-term capacity development Given current evaluation capacity within all parts of the international development sector, this should be viewed as a medium- to long-term undertaking that will not have immediate results. Oxfam GB's recent reflection on its first effectiveness study, which concluded it was overambitious, is a lesson to us all:

'Worth noting, then, that even a large NGO like Oxfam has serious limits to the capacity we have available for impact evaluation. The silver bullet is to get an appropriate balance of investment versus delivery, and to ensure that external accountability purposes do not outweigh the potential to use this evidence for learning and improvement' (Jennie Richmond, Head of Programme Performance and Accountability, Oxfam GB).¹⁰

Develop more critical epistemological awareness

Chambers (1998) coined the phrase 'self-critical epistemological awareness', meaning being aware of how our backgrounds, and the power relations that shape our

positionality, affect our interpretations and constructions of knowledge. Although there has been some progress, power relations between aid players (from the grassroots to institutional funders) mean that admission of weakness or failure is still rare and methodological discussions are undervalued or weak. There is a need for self-critical epistemological awareness to spread so that INGO actors can nurture organisational cultures that celebrate learning from admitting challenges or weak practice.

Attitudinal and cultural change Developing NGO capacity to do better M&E and impact assessment and real-time learning is about more than developing technical skills; it will require attitudinal change as well. New 'quality of evidence' tools (Bond n.d.) emerging in the sector suggest a willingness to borrow and learn from practices more common in applied academic research, including the routine presentation of a critical discussion of one's own methodology. While this is virtually an indicator of quality in academic research, it is rarely seen in NGO impact assessments or evaluations.

These implications should not be perceived as yet more top-down requirements that make it harder still for country-based programme staff and partners to report to headquarters. In many respects, because they invite realism into the practice of M&E and impact assessment, they can reduce the tension often felt by country-based staff and partners between the requirement to comply with standardising, homogenising results-focused M&E requirements, and the need to acquit themselves and their programmes in ways that do justice to the complex and not entirely manageable realities surrounding them.

To some extent, INGO efforts to make the shifts highlighted above are hindered by the tools used by results-based management-focused donors – for instance, M&E systems driven by logical framework indicators. In some instances they are also constrained by decisions the INGOs have made about their own internal performance management systems – for example the use of global indicators (Hughes and Hutchings 2011).

However, among the evaluations we reviewed we found some evidence of change in both INGOs and donor agencies. INGOs increasingly recognise the weaknesses of 'projectable' approaches to evaluation driven by logframe indicators described in terms of effects on numbers of individual people. Similarly, evaluation staff in donor agencies like DFID are starting to implement the recommendations of related reviews that they have commissioned (e.g. Stern *et al.* 2012), developing frameworks to identify 'difficult-to-measure' programmes and encouraging NGOs to take more adaptive learning approaches.¹¹

The scoping we undertook revealed some interesting designs that indicate the four INGOs are (within some of

Box 2 Signs of INGO responsiveness to challenges in M&E of E&A programmes

CARE Bangladesh is piloting an innovative participatory approach to monitoring and evaluating a livelihoods programme. The main objective of the M&E system is community learning to improve programme effectiveness, but it integrates a statistical analysis of changes in household assets and income with quantitative analysis of data gathered from a 'Most Significant Change'¹² approach to assessing various dimensions of empowerment, some resulting from enhanced government accountability.

CARE Peru integrated a quantitative and qualitative analysis of changes resulting from a community health surveillance system and integrated it with a participatory action research project that explored how women involved in the project experienced empowerment in public and private spheres.

World Vision Uganda used quantitative and qualitative methods to pilot test a 'theory of change'-based evaluation methodology to assess increases in knowledge of rights and perceptions of influence over decisions, in a Citizen Voice and Accountability project. This evaluation benefits from its attempts to link learning from assessment of results to the quality of project implementation, particularly the knowledge and capacity of NGO staff, which is important but rare in INGO evaluation practice.

Plan UK was able to produce quantitative data on increases in access to services across many countries that resulted from DFID PPA funding. Partly as a result of criticisms from PPA evaluators that there were not enough qualitative data to explain the meaning of quantitative results, it is now doing more work on qualitative analysis including using tools such as outcome harvesting.

The Mid-Term Review of **Christian Aid's** Governance and Transparency Fund programme, Power to the People, introduced a generic framework for understanding E&A change in a meta-evaluation of heterogeneous projects being implemented in diverse contexts.

their programmes at least) actively responding to the issues above (see Box 2). However, given the size and complexity of the four organisations, they face extreme time, resource and capacity constraints for pulling guidelines, recommendations and lessons together so that they can be implemented consistently in projects carried out in very different contexts.

Within these studies the NGOs demonstrated examples of good practice and some capacity to measure changes in:

- knowledge of rights;
- perceptions of relationships between different groups;
- skills acquired;
- issues solved; and
- access to infrastructure and services resulting from engagement with state actors.

In some instances there was validation of these changes by communities. However, there were weaknesses in overall analysis – particularly of qualitative data that was found to be very time-consuming – and efforts to explore causal mechanisms. Outcomes from theories of change were used for descriptive rather than explanatory purposes. In other words, efforts to tease out how and why changes have occurred and the relative effects of project interventions compared to other contextual factors could do with further improvement. It is quite possible that the issues identified are already beginning to be addressed. Our scoping revealed increasing investment in baselines, some informed by detailed context analysis; the establishment of monitoring systems to support evaluation and impact assessment earlier in programmes; and the development of participatory indicators and scalar tools to aid measurement. Plan and World Vision are testing theory-based ToC approaches to evaluation. Some specific examples of current INGO E&A evaluation initiatives are referred to in appropriate places in the guidelines produced from the scoping exercise, summarised below and published in full as CDI Practice Paper 01 Annex.

4 Guidelines to improve the evaluability of E&A programmes

This section summarises guidance that can be followed when designing MEL strategies and systems capable of assessing the outcomes and impacts of E&A programmes and exploring causal mechanisms. As this is an iterative and non-scientific process, the guidance needs to be applied differently to programmes according to the values of organisations, strategic importance of the respective programme, framing of evaluation objectives and questions, resources available, programme attributes and the context of implementation. Set out as a list of pointers to think about corresponding to stages in the process, they are intended to assist UK- or country-based programme managers, M&E staff, those responsible for commissioning evaluations and impact assessments, and those who write funding proposals to donors. Some pointers may suggest the need for external assistance, and may prove useful in developing terms of reference for this.

When deciding when to start work on designing a MEL, impact evaluation or impact assessment system:

- Start early to ensure that opportunities are maximised for MEL strategies to contribute to real-time learning

to improve the impact and value for money of programmes, evaluations and impact assessment.

When deciding on MEL designs, bear in mind:

- INGO organisational values and norms favour participatory design elements;
- The current state of evidence on E&A impact and the complex, unpredictable and long-term nature of E&A programmes mean experimental and quasi-experimental approaches, which are expensive, have limited utility as single designs or indeed as hybrids;
- Theory-based approaches guided by ToCs are the most appropriate dominant design, though they can be complemented by participatory elements, surveys or case studies;
- Design decisions should be influenced by how important it is to infer causality as compared to understanding causal mechanisms and dynamics in any given case;
- Considering particular programme attributes and the contexts in which they are being implemented helps to identify specific evaluation challenges and implications for design;¹³
- Resource and capability constraints place pragmatic limitations on the scope of MEL strategies and need to be considered early on.

When operationalising a MEL strategy:

- Operationalising MEL designs requires choosing which outputs, milestones towards outcomes and – if appropriate – impacts will be identified and assessed in a given timeframe;
- Decisions need to be taken about trade-offs between local utility and standardisation for donors;
- Sample designs need to take into account demographics and power relations at different levels;
- Choices of measurement tools need to be guided by programme attributes and consideration of how different tools and triangulation approaches are likely to enable interrogation of ToC assumptions;
- Quality is often more important than quantity when thinking about indicators, sample designs and measurement tools;
- Check to ensure statistical analysis techniques for ordinal scoring data are valid if they are being considered;
- Make sure qualitative approaches consider the associated costs of translation and analysis.

When implementing the programme and the MEL strategy:

- Plan an inception workshop to iron out the details of the MEL operational plan, including roles and responsibilities;
- Develop protocols with adequate capacity development for those involved in implementing the plan;
- In addition to planning a final evaluation, ensure resources are available for periodically revisiting ToCs, interrogating assumptions, analysing contributions with reference to context, and adapting programmes in the light of learning.

When reviewing or writing MEL reports or writing self-critical appraisal of own methodology:

- Make evident how, when and why evaluation decisions were taken, for example design, sample size, approach to analysis;
- Discuss limitations, doubts, and the positionality or bias of the researchers;
- Ensure conclusions use findings to assess the validity of assumptions and ToCs, and discuss the implications for the programme in question and other practitioners.

5 Conclusion

The review found that the four INGOs engaged in this project are already improving the evaluability of empowerment and accountability programmes in order to enhance accountability and learning at different levels in complex aid relationships.

The very nature and complexity of E&A programmes together with the relatively weak evaluation capacity within the international development sector suggests this is going to be a long-term project. Improving the evaluability and assessment of the outcomes of such programmes in large, complex, international organisations trying to decentralise power and work in partnership to shift rather than entrench power inequalities, is an extremely difficult task.

Getting it 'right' – or righter – is going to require investments of time, attention and funding in the short to medium term, to shift the current INGO organisational capacity, organisational self-expectations, and expectations of institutional donors operating in an evidence-based paradigm, onto a different level. It requires taking a more integrated approach – informed by complexity science – to evaluation strategy design, in which final evaluations or impact assessments are no longer discrete activities, but part of a longer learning process. Such holistic designs may call for radical rethinking of 'independent evaluation' and relationships between consultant evaluators and insiders. For example, contracting the same consultants to help with initial designs as with final evaluations – an approach some donors are piloting – could improve the quality of analysis and final evaluation.

Given that resources are constrained, strategic thinking about how and where to invest these are urgent on the INGOs' side, and realistic and strategic adjustments of expectations would be welcome on the institutional donors' side. As many donors are realising, funders anxious to demonstrate the effectiveness of their spending do nobody any favours by pushing INGOs into producing data that lack meaning. They do everyone favours by working with these INGOs to resource their MEL capacity strategically, proportionately to need, and in keeping with cost-benefit considerations. One related question arising from recent literature on impact evaluation is whether

donors should be asking INGOs implementing 'demand-side' empowerment and accountability programmes to evaluate their 'impact' on the lives of individuals independently of complementary supply-side governance capacity-building initiatives. Approaches that 'nest' INGO evaluations within broader donor governance portfolios that value and aim to explore causal links between INGO contributions and the development of diverse civil societies, rather than encouraging them to make tenuous claims in terms of *their* impacts on the lives of individual people, may be a more appropriate strategy.

Notes

- 1 Reference is made in this paper, especially Section 3, to some of these examples reviewed. Their full bibliographic details are provided in the companion CDI Practice Paper 01 Annex. Some are internal INGO documents that are not publicly available.
- 2 From the terms of reference of the Stern *et al.* (2012) study.
- 3 The Transparency and Accountability Initiative (T/A Initiative) is a donor collaborative committed to strengthening democracy and development through empowering citizens to hold their governing institutions to account. Its members are DFID, Ford Foundation, Hivos, International Budget Partnership, Omidyar Network, Open Society Institute, Revenue Watch Institute, and William and Flora Hewlett Foundation. For the outputs of the mentioned review, see www.transparency-initiative.org/workstream/impact-learning.
- 4 Shutt and McGee (2012).
- 5 Examples of how some of these designs have been used in E&A programmes can be found in the IETA review by McGee and Gaventa (2010).

References

- Barder, O. and Ramalingam, B. (2012) *Complexity, Adaptation and Results*, blogpost, <http://blogs.cgdev.org/globaldevelopment/2012/09/complexity-and-results.php> (accessed 1 November 2012)
- Bond (n.d.) *An Introduction to the Principles for Assessing the Quality of Evidence*, www.bond.org.uk/data/files/Effectiveness_Programme/120828Full_Bond_checklist_and_guide.pdf (accessed 6 March 2013)
- Chambers, R. (1983) *Rural Development: Putting the Last First*, UK: Longman
- Chambers, R. (1998) 'Beyond "Whose Reality Counts?": New Methods We Now Need?', *Studies in Cultures, Organizations and Societies* 4.2: 279–301
- Davies, R. and Dart, J. (2005) *The Most Significant Change Technique: A Guide to its Use*, www.mande.co.uk/docs/MSCGuide.pdf (accessed 27 February 2013)
- Hughes, K. and Hutchings, C. (2011) *Can We Obtain the Required Rigour Without Randomisation? Oxfam GB's Non-experimental Global Performance Framework*, Working Paper 13, International Initiative for Impact Evaluation, www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_13.pdf (accessed 28 February 2013)
- McGee, R. and Gaventa, J. (2011) *Shifting Power: Assessing the Impact of Transparency and Accountability Initiatives*, IDS Working Paper 383, Brighton: Institute of Development Studies
- McGee, R. and Gaventa, J. (2010) *Synthesis Report: Review of Impact and Effectiveness of Transparency and Accountability Initiatives*, Institute of Development Studies, Brighton, www.transparency-initiative.org/workstream/impact-learning (accessed 18 February 2013)
- Pauson, R. (2002) 'Realist Synthesis: Supplementary Reading 2 – "In Search of a Method"', *Evaluation* 8.2: 157–81
- Porter, D.; Bryant, A. and Thompson, G. (1991) *Development in Practice: Paved with Good Intentions*, London: Routledge
- Ramalingam, B. and Jones, H. with T. Reba and J. Young (2008) *Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts*, ODI Working Paper 285, London: Overseas Development Institute
- Reeler, D. (2007) *A Three-fold Theory of Social Change and Implications for Practice, Planning, Monitoring and Evaluation*, Centre for Development Practice, www.cdpra.org.za/articles/A%20Theory%20of%20Social%20Change%20by%20Doug%20Reeler.pdf (accessed 21 November 2012)
- Roche, C. and Kelly, L. (2012) *The Evaluation of Politics and the Politics of Evaluation, DLP Background Paper 11, Developmental Leadership Program*, summary available at www.gsdr.org/go/display&type=Document&id=4340 (accessed 27 February 2013)
- Rondinelli, D. (1993) *Development Projects as Policy Experiments: An Adaptive Approach to Development Administration*, London: Routledge
- Shutt, C. (2010) 'Monitoring, Evaluating and Assessing the Impact of Governance Programmes: Summary of Literature and Practice Review', paper commissioned by CARE, <http://governance.care2share.wikispaces.net/file/view/Governance+M%26E+Literature+Review+Final+6.10.11.pdf> (accessed 18 February 2013)
- Shutt, C. and McGee, R. (2013) CDI Practice Paper 01 Annex, March, Brighton: Institute of Development Studies
- Shutt, C. and McGee, R. (2012) 'Improving the Evaluability of INGO Empowerment and Accountability Programmes', unpublished paper for CARE, Christian Aid, Plan UK and World Vision UK, November
- Stern, E.; Stame, N.; Mayne, J.; Forss, K.; Davies, R. and Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, London: Department for International Development
- York, N. and Hoy, C. (2012) *What do DFID Wonks Think of Oxfam's Attempt to Measure its Effectiveness?*, www.oxfamblogs.org/fp2p/?p=12254 (accessed 27 February 2013)

These recommendations offer neither prescriptions nor magic bullets. They should be seen as a contribution to ongoing discussions amongst practitioners who are working hard to enable their agencies to respond to new evaluation challenges. By engaging with evaluation dilemmas early on in programme conceptualisation it should be possible to select designs and methods that are cost-effective and appropriate for assessing change given the strategic importance of the interventions; their attributes and theories of change underpinning them; the resources and capabilities available; and organisational values.

- 6 Further details can be found in Shutt (2010).
- 7 For example, Chambers's *Rural Development: Putting the Last First* (1983), Rondinelli's *Development Projects as Policy Experiments: An Adaptive Approach to Development Administration* (1993) and Porter *et al.*'s *Development in Practice: Paved with Good Intentions* (1991).
- 8 www.oxfamblogs.org/fp2p/?p=12254 (accessed 27 February 2013).
- 9 *Ibid.*
- 10 <http://policy-practice.oxfam.org.uk/blog/2012/10/effectiveness-reviews> (accessed 27 February 2013).
- 11 DFID Investment Committee Paper: *Workstream on Harder to Measure Benefits and Results*, Discussion Note for meeting, 15 November 2012.
- 12 See Davies and Dart (2005).
- 13 A tool can be found in the companion CDI Practice Paper 01 Annex.

“ If international NGOs are to successfully ‘measure’ or assess outcomes and impacts of empowerment and accountability programmes, they need to shift attention from tools and methods to developing more holistic and complexity-informed evaluation strategies during programme design. This will require investments of time, attention and funding in the short to medium term, to shift the current INGO organisational capacity, organisational self-expectations, and expectations of institutional donors operating in an evidence-based paradigm, onto a different level. ”

Centre for Development Impact (CDI)

The Centre is a collaboration between IDS (www.ids.ac.uk) and ITAD (www.itad.com).

The Centre aims to contribute to innovation and excellence in the areas of impact assessment, evaluation and learning in development. The Centre’s work is presently focused on:

- (1) Exploring a broader range of evaluation designs and methods, and approaches to causal inference.
- (2) Designing appropriate ways to assess the impact of complex interventions in challenging contexts.
- (3) Better understanding the political dynamics and other factors in the evaluation process, including the use of evaluation evidence.

This CDI Practice Paper was written by **Cathy Shutt** and **Rosie McGee**. The authors would like to thank Jake Allen (Christian Aid), Jennifer Doherty (Plan UK), Nicole Walshe (CARE UK), Gaia Gozzo (CARE UK) and Daniel Stevens (World Vision UK) for helping shape this output. We are also grateful to Jake Phelan (Plan UK), Hilary Williams (World Vision UK), John Lakeman (CARE UK), Francesca Scott (Christian Aid), Ariel Frisncho Arroyo (CARE Peru), Donald Mogeni (World Vision UK) and Bill Walker (World Vision Australia) for sharing their time, experience and ideas that improved the report.

The briefing is based on ‘**Improving the Evaluability of INGO Empowerment and Accountability Programmes**’, an unpublished paper by **Cathy Shutt and Rosie McGee** (2012).

The opinions expressed are those of the authors and do not necessarily reflect the views of IDS or any of the institutions involved. Readers are encouraged to quote and reproduce material from issues of CDI Practice Papers in their own publication. In return, IDS requests due acknowledgement and quotes to be referenced as above.

© Institute of Development Studies, 2013