



AUSTRALIAN
COUNCIL
FOR
INTERNATIONAL
DEVELOPMENT



RESEARCH FOR
DEVELOPMENT
IMPACT NETWORK

A collaboration between
the Australian Council for
International Development
and Australian universities

Demonstrating Outcomes and Impact across Different Scales

RESEARCH REPORT

Jo Hall
June 2017

About ACFID

The Australian Council for International Development (ACFID) is the peak body for Australian non-government organisations (NGOs) involved in international development and humanitarian action. Our vision is of a world where all people are free from extreme poverty, injustice and inequality and where the earth's finite resources are managed sustainably. Our purpose is to lead and unite our members in action for a just, equitable and sustainable world.

Founded in 1965, ACFID currently has 125 members and 21 affiliates operating in more than 95 developing countries. The total revenue raised by ACFID's membership from all sources amounts to \$1.9 billion (2016), \$930 million of which is raised from over 1.5 million Australians (2016). ACFID's members range between large Australian multi-sectoral organisations that are linked to international federations of NGOs, to agencies with specialised thematic expertise, and smaller community based groups, with a mix of secular and faith based organisations.

ACFID members must comply with the ACFID Code of Conduct, a voluntary, self-regulatory sector code of good practice that aims to improve international development and humanitarian action outcomes and increase stakeholder trust by enhancing the transparency, accountability and effectiveness of signatory organisations. Covering 9 Quality Principles, 32 Commitments and 90 compliance indicators, the Code sets good standards for program effectiveness, fundraising, governance and financial reporting. Compliance includes annual reporting and checks. The Code has an independent complaints handling process.

Further information can be found at www.acfid.asn.au.

About the Research for Development Impact Network

The Research for Development Impact (RDI) Network is a collaboration between the Australian Council for International Development (ACFID) and the Australian universities. It is a network of practitioners, researchers and evaluators working in international development with the objective of linking quality research, policy and practice for impact in international development.

The Network began in 2009 and grew out of a collective desire to widen debate on international development and to strengthen collaboration between academics and members of ACFID. Since this time, the Network has continued to grow and promote positive relationships and connections between ACFID members and universities, with the overall goal of supporting collaboration and understanding across actors within the Australian development sector.

Further information can be found at www.rdinetwork.org.au.

Acknowledgements

This research report was authored by Jo Hall, independent consultant, and commissioned by the Development Practice Committee (DPC) of ACFID and the RDI Network.

ACFID acknowledges the input of its member agencies through participation in the research. ACFID particularly acknowledges the contribution of the Department of Foreign Affairs and Trade (DFAT) as a part of the steering committee supporting this research.

© Australian Council for International Development 2017

Table of Contents

1	Introduction	1
2	Context.....	2
2.1	The changing landscape of aid and development effectiveness	2
2.2	Complexity.....	4
2.3	Fragmentation	5
2.4	Monitoring and Evaluation	5
2.5	The challenge for NGOs: measuring at scale.....	7
2.6	Factors to consider when capturing, integrating and reporting on outcomes and impact across different scales.....	10
3	Methods and approaches for capturing, integrating and reporting on outcomes and impact across different scales	11
3.1	Using indicators	13
3.1.1	Aggregating results	13
3.1.2	Framing indicators	13
3.1.3	Coalition indicators.....	14
3.1.4	Rating indicators	14
3.1.5	Strengths and limitations of using indicators.....	15
3.2	Evaluative approaches.....	17
3.2.1	Strategic evaluation	18
3.2.2	Strengths and limitations of strategic evaluation.....	18
3.2.3	Evaluation synthesis	21
3.2.4	Strengths and limitations of evaluation synthesis	23
3.3	Research approaches	24
3.3.1	Case studies	25
3.3.2	Qualitative methods	25
3.3.3	Strengths and limitations of case study and other qualitative methods.....	26
3.4	Combining the methods	28
3.4.1	Donor-led program approaches.....	28
3.4.2	Agency-led approaches	28
3.4.3	Considerations in combining methods	29
4	Conclusions and Recommendations.....	31
4.1.1	Being clear on the purpose and questions and making use of the information.....	32
4.1.2	Applying fit-for-purpose methods in fit-for-purpose ways.....	33

4.1.3	Adopting methods and approaches that address complexity	34
4.1.4	Considering the needs of all partners, including locally	34
4.1.5	Capturing the distinctive contribution of the NGOs	35
4.1.6	Not being overly complicated, technocratic or exacerbating fragmentation	35
4.1.7	Having adequate resourcing to meet the purpose	36
4.2	Recommendations	36
Annex 1: Research Methodology Summary.....		38
Annex 2: Istanbul Principles for CSO Development Effectiveness.....		40
Annex 3: Case Studies		41
CASE STUDY 1: Tracking Impact: An exploratory study of the wider effects of Norwegian civil society support to countries in the South.....		41
CASE STUDY 2: Reach, Relationships and Results: Case Studies of Australian NGOs' work in Education		44
CASE STUDY 3: Exploring the link between child and youth participation and development effectiveness		46
CASE STUDY 4: Australia Africa Community Engagement Scheme		49
CASE STUDY 5: OXFAM Australia: Strategic Plan (2014-19) Reporting		54
References.....		58

1 Introduction

ACFID is committed to enabling its members to *improve, prove and demonstrate their individual and collective effectiveness and impact*. ACFID has and will continue to support its members to improve their monitoring and evaluation practices, including through setting standards in the ACFID Code, brokering access to training, curating monitoring and evaluation resources distributed through ACFID's on-line Good Practice Toolkit and Resource Library and convening structured conversations on emerging issues in monitoring and evaluation. Until now, ACFID has not supported agencies to capture, integrate and report on outcomes and impact across different scales. The purpose of this research is to show how evidence of outcomes and impact can be better captured, integrated and reported on across different scales of work.

It is the first piece of a larger body of work over the next few years, when ACFID and the Research for Development Impact (RDI) Network will – resources permitting – use the research findings to inform the development of other pieces of work such as a user guide for practitioners. The report is targeted to the ACFID membership, particularly to monitoring, evaluation and learning (MEL)¹ specialist staff in Australian NGOs working in international development. It is also targeted to members of the RDI Network, including monitoring and evaluation consultants and academics working with Australian NGOs. It may also be useful for other NGOs, partner organisations and donors, particularly DFAT.

The research was guided by a Steering Committee² and adopted a pragmatic and participatory approach, including five case studies, a review of the current literature, a workshop with monitoring and evaluation practitioners from Australian NGOs and other stakeholders, feedback from academics and NGOs (including the Development Practice Committee of ACFID and the RDI Network) on a working paper, and the analysis of all findings in preparing this report (see Annex 1 for a summary of the methodology). It addresses the following questions:

1. What are the different scales of work that are relevant to this study from the perspective of ACFID member agencies, ACFID and DFAT?
2. Which factors should inform the selection, application and critical assessment of practices for capturing, integrating and reporting on evidence of outcomes and impact across these scales?
3. What approaches, methods and tools are available to capture, integrate and report on evidence of outcomes and impact across these scales?
4. Taking account of the principles referred to above, what are the strengths and weaknesses of each of these approaches, methods and tools when applied at these scales?
5. What issues emerge – both political and technical - when applying these approaches, methods and tools at these scales and how can they be addressed?
6. What factors are helping/hindering the take up and principled use of these approaches, methods and tools at these scales by individuals, by agencies, by groups of agencies and by donors?

¹ Throughout this report “monitoring and evaluation” or “M&E” is used as the universal term for activities associated with tracking and reporting on results, outcomes and impact. “Monitoring, evaluation and learning” or “MEL” is the preferred term in the Australian and UK NGO community, and is used in this report to emphasise the learning and implementation aspect of undertaking monitoring and evaluation.

² Phil Lindsay from ACFID's Development Practice Committee, Juliet Willetts from the Research for Development Impact Network, Heather Fitt from DFAT and Chris Adams from ACFID.

7. What can agencies, ACFID, RDI Network and DFAT do to improve performance in this area?

This report is in three chapters. The Context chapter addresses the first two research questions, the Methods and Approaches chapter addresses research questions 3 to 6, and the Conclusions and Recommendations chapter makes four recommendations for ACFID as an initial response to question 7.

2 Context

It is important to locate this research in the current development context, because much has changed over the last twenty years, including the way different stakeholders interpret development effectiveness, complexity and fragmentation issues and the ways that monitoring and evaluation have evolved. All of these have a direct bearing on the research and on how ACFID members might consider the different ways that evidence of outcomes and impact can be better captured, integrated and reported on across different scales.

2.1 The changing landscape of aid and development effectiveness

The aid 'landscape' has dramatically changed in recent years, with agreements on truly global goals and a more inclusive and proportional understanding of aid. For the first time, the closest thing to a global consensus on what success looks like has been achieved through the global agreement on seventeen Sustainable Development Goals (SDGs), effective 1 January 2016.

The Fourth High Level Forum on Aid Effectiveness, a global conference held in Busan, South Korea in 2011, strongly influenced a reshaping of the way we think about how aid and development effectiveness is defined. Earlier aid effectiveness conferences had established that country ownership and common development goals were central to aid effectiveness, but Busan formally opened the field to include the aid-like assistance of the 'emerging' or 'new' development partners (such as Brazil, China and the Gulf countries) and the role of civil society³ (including NGOs) as well as the Organisation for Economic Co-operation and Development (OECD) donors (Mawdsley et al., 2014). Notwithstanding these global advances, the role of civil society in some countries has been increasingly constrained in terms of foreign funding (Christensen and Weinstein, 2013).

Another important change has been a shift in the language itself, from aid to development cooperation in recognition that aid per se is small compared to foreign direct investment and remittances from overseas workers, and that a number of 'southern' countries do not view their cooperation with developing countries as aid (Gore, 2013). Nonetheless, aid-like assistance deliberately targeting development from OECD DAC donors, non-OECD DAC donors and philanthropic giving is significant, totalling about 200 billion US Dollars in 2011 (Center for Global Prosperity, 2013, Greenhill and Prizzon, 2012).

The language also changed from aid effectiveness to development effectiveness, but with mixed interpretations about what this actually means. Capturing, integrating and reporting on

³ Civil society is defined for this report as the arena where people organise around and deliberate on shared collective purposes. Civil society is often populated by organisations that vary in their degree of formality and typically includes associational forms such as trade unions, social movements, developmental non-government organisations, virtual networks, campaigns, coalitions, faith groups, think tanks, research institutions, direct action groups, peace groups, human rights organisations and, in some cases, clan groups (Howell and Hall, 2012). Some civil society groups work against the interests of development, for example in financing terrorism, but for the purpose of this report it is 'liberal' civil society to which it refers.

outcomes and impact across different scales in this research is about a more strategic approach to measuring development effectiveness by Australian NGOs.⁴ But they need to communicate this in an environment of multiple perspectives. For example, the 'emerging' donors think of development cooperation (and effectiveness) in terms of mutual benefit and non-interference, principles established in 1955 at the Asia Africa Conference held in Bandung, Indonesia. The five aid effectiveness principles decided at the Second High Level Forum on Aid Effectiveness held in Paris (see box 1) were agreed among the OECD donors in 2005. In 2008, the OECD donors agreed to accelerate action on these principles at the Third High Level Forum in Accra, and also to bring the 'emerging donors' and civil society into the fold.

Box 1: The Five Principles of the Paris Declaration on Aid Effectiveness (OECD)

- 1. Ownership:** *Developing countries set their own strategies for poverty reduction, improve their institutions and tackle corruption.*
- 2. Alignment:** *Donor countries align behind these objectives and use local systems.*
- 3. Harmonisation:** *Donor countries coordinate, simplify procedures and share information to avoid duplication.*
- 4. Results:** *Developing countries and donors shift focus to development results and results get measured.*
- 5. Mutual accountability:** *Donors and partners are accountable for development results.*

The language of 'development effectiveness' (as opposed to 'aid effectiveness') was introduced in the lead-up to the Accra meeting (OECD, 2007). However, there are multiple lines of thought about what this actually means (Kim and Lee, 2013), including:

- a focus on gender, the environment and human rights as more intrinsic to development which was otherwise dominated by economic growth (Eyben, 2013);
- a focus on the outcomes of aid interventions (Kindornay, 2011)
- a focus on broader development outcomes regardless of what brought them about (Kindornay, 2011)
- the 'beyond aid' agenda which includes broader development finance and private sector flows, as well as policy coherence across donor governments, not just in aid, but in policies like migration and trade (Eyben, 2013, Kindornay, 2011, Mawdsley et al., 2014).

ACFID's interpretation of development effectiveness is most closely associated with the outcomes of aid interventions, but also includes elements of development cooperation targeting root causes of poverty and marginalisation. The ACFID definition of development effectiveness is:

'Promoting sustainable change which addresses the causes as well as the symptoms of poverty and marginalisation – i.e. reduces poverty and builds capacity within communities, civil society and government to address their own development priorities.' (ACFID, 2015).

This interpretation arises from a parallel process, following Accra, when civil society organisations developed a global consensus on development effectiveness for Civil Society Organisations (CSOs). In Istanbul (2010), the CSOs developed eight principles for CSO

⁴ Throughout this report, Australian NGOs or ANGOs is used to refer to NGOs based in Australia, working in the international development sector. Specifically, the focus is on NGOs who are members of ACFID.

development effectiveness (see Annex 2) that were later developed into an International Framework for CSO Development Effectiveness in Cambodia in 2011 (Open Forum for CSO Development Effectiveness, 2011).

Capturing, integrating and reporting on outcomes and impact across different scales for the Australian NGOs needs to be consistent with the shifting aid and development landscape, as reflected in academic literature as well as the international commitments outlined above. This includes promoting sustainable change which addresses the causes as well as the symptoms of poverty and marginalisation – i.e. reducing poverty and building capacity within communities, civil society and government to address their own development priorities.

2.2 Complexity

Another characteristic of the changing landscape is an increasing recognition of complexity. Aid interventions are usually complex and they operate in complex, and changing, contexts. The way in which the implementers and recipients interpret and respond to the intervention, their relationships with each other, the institutional setting of customs, rules and norms and the broader social, economic and cultural environment in which the intervention operates all interact in a complex web to produce the ‘outcomes’ of the intervention (Pawson, 2014). For many years, economic growth and development was considered a universal and linear process that could be accelerated by an injection of capital or a change in policy, but these ideas are now challenged by complexity theories in the social sciences. In his aptly titled book ‘Aid on the Edge of Chaos’, Ramalingam claims ‘serious mistakes in all walks of public and political life are increasingly being attributed towards seeing interconnected, messy problems as simple, closed controllable ones.’ (Ramalingam, 2013: 130). Understanding the complexity of the particular context in which an aid intervention is operating as well as the possibility of *emergent* outcomes (when the interaction of the various parts of the complex system itself give rise to changes) are now important considerations in measuring the effectiveness of aid interventions (Rogers, 2008). The implications for the way that NGOs work are significant. Duncan Green, Senior Strategic Adviser for OXFAM GB urges NGOs to acknowledge complexity and change practice accordingly, including addressing the ways that performance is measured (Green, 2015).

In recognising complexity, the need to think and work politically has also emerged as a relatively new theme over the past ten years. It is about shifting mindsets to acknowledge politics as one crucial element of development, along with a focus on the roles of agency, leadership and coalitions in bringing about change within a deeper understanding of the ‘micro-politics’ at all levels of a particular context (Leftwich, 2011). Working politically in development means:

... supporting, brokering, facilitating and aiding the emergence and practices of developmental or reform leaderships, organizations, networks and coalitions, in the public and private fields, at all levels, and across all sectors, in response to, and in concert with, initiatives and requests from local individuals and groups (Leftwich, 2011: 8)

The main characteristics of what this looks like in practice have implications for the way NGOs think about their work and their M&E systems. Those characteristics include:

- Developing a compelling narrative
- Adopting flexible and adaptable approaches
- Developing greater risk tolerance
- Identifying and supporting the right staff
- Brokering spaces for enhanced collaboration (Menocal, 2014)

More flexibility and adaptability means fewer pre-defined results and mixed approaches to monitoring and evaluation that take account of both the politics of development and the politics of evaluation (Roche and Kelly, 2012). Enhanced collaboration implies a need to measure the performance of collaborations and partnerships and the added value they bring to achieving outcomes and impact.

There are other approaches to addressing complexity in monitoring and evaluation - this report recognises that it is a relatively new field and there are multiple and emerging ideas. One approach for example includes acknowledging that interventions work differently in different contexts. This is the basis of realist evaluation and realist synthesis, discussed in the next chapter.

Emerging understanding of complexity needs to be accounted for when considering methods and approaches used to capture, integrate and report on outcomes and impact across different scales (see section 4.1.3).

2.3 Fragmentation

A final and related element of the changing aid landscape has been the increasing fragmentation of aid, adding complicatedness to its complexity as well as increasing the administrative burden on developing countries and decreasing effectiveness. Globally 'there has been an extraordinary explosion in the number of state and non-state aid actors and programmes' (Mawdsley et al., 2014). In the early 2000s, Vietnam, for example, hosted 25 official bilateral aid donors, 19 multilateral agencies and over 350 NGOs operating 8,000 aid projects (Acharya et al., 2006). In 2007, aid projects globally numbered over 90,000, with fragmentation (more numbers of smaller projects) increasingly severe in the social sectors (Frot and Santiso, 2009). Solving the problems of fragmentation has not been easy for the donors or for NGOs; a review of fragmentation among 115 Belgian NGOs suggests that the incentive structure for the NGOs limits their ability to address the issue (Molenaers et al., 2014). The internal political drivers of behaviour in international NGOs (Lissner, 1977) are often at odds with the changes needed to address issues like complexity and fragmentation. One of the many consequences of increasing fragmentation has been that it is becoming increasingly difficult to get a clear picture of what it is all 'adding up to', discussed below in section 2.5.

2.4 Monitoring and Evaluation

Monitoring and evaluation of aid efforts have always been important but the emphasis on results has increased considerably over recent years, driven in part by the Paris 'results' agenda. Large investments have been made and the term 'M&E' has become part of the development lexicon with a large body of expert practitioners and consultants now populating the already crowded field of actors. However, the promise of deepening the global understanding of what works and what doesn't or of helping partner countries select likely successes in their ownership of development programs have not yet been realised. M&E has been subject to the same issues of complexity, fragmentation, differing world views and politics as the rest of aid and development.

Broad critiques of donor efforts in monitoring and evaluation approaches and results-based management range from the over-simplification of program logics, measuring things that are measurable rather than meaningful (Ramalingam, 2013, Natsios, 2010), the need for a more evaluative culture in donor institutions (Mayne, 2009) to concerns about the mechanical use of performance information for reasons of accountability and control (Vähämäki et al., 2011, Eyben, 2015). Compounding these issues is the fact that 'results' for some mean development outcomes rather than how any intervention has contributed towards the achievement of those development outcomes. Another issue has been that often the outputs of the M&E efforts are not useful or used by donors, NGOs or partner countries. The over-bureaucratisation of aid and the 'rising tide of technocracy that has swept through the world

of foreign aid over the last 10 years' (Banks et al., 2015: 708) has extended to its monitoring and evaluation systems.

The Paris principles promote country-owned monitoring and evaluation systems which should help inform countries as to which projects are working well with a view to assisting them in selecting and implementing aid projects. In practice, the efforts have tended to be overly technical at the expense of generating meaningful information.

A case study of Rwanda's monitoring and evaluation system is typical. It points to fragmentation in monitoring and evaluation between donors and criticises monitoring efforts as overly technocratic and lacking evaluative analysis (Holvoet and Rombouts, 2008). Increased numbers of parallel monitoring and evaluation systems established for increased numbers of actors and programs risk an imposed burden on the developing country without a corresponding benefit. The inability of donors to harmonise with each other or to allow partner countries to take the lead means there are likely to be a proliferation of systems generating information for donors rather than a coordinated investment focused on strengthening partner systems to generate meaningful development information for their own needs.

NGOs have invested heavily in monitoring, evaluation and learnings systems. As the logical framework has diminished in popularity (with many donor agencies, but not all) and theories of change - more aligned with the complexity of aid and development - become more commonplace, the sophistication of monitoring, evaluation and learning has developed to some degree in areas like measuring empowerment or advocacy efforts or partner satisfaction surveys. Numbers of project evaluations have dramatically increased over the past 20 years, guided by the OECD evaluation criteria of relevance, effectiveness, efficiency, impact and sustainability (OECD/DAC). Evaluation methods have evolved rapidly and evaluation is on its way to becoming an academic discipline in its own right. Nonetheless, practice is mixed, and in particular the need to address complexity in M&E systems described earlier has not yet changed the majority of practice.

While international development NGOs have been involved in helping both governments and NGOs develop their own civil society regulatory systems, they have been more focused on strengthening their own monitoring and evaluation systems than those of their in-country partners. Similarly the international NGOs have been less concerned with strengthening the capacity of civil society to monitor government performance and achievement of the SDGs. This seems inconsistent with the Australian NGO commitment to build capacity within communities, civil society and government to address their own development priorities.

Some academics, however, have been occupied with issues of downward accountability, to address the need for NGOs to be as accountable to beneficiaries as they are to institutional donors (Roche, 2009, O'Dwyer and Unerman, 2008, Schmitz et al., 2012, Kilby, 2010). NGO practice in this area seems less advanced than the rhetoric, however. A study of some of the complexities involved are uncovered in a review of NGO support for the Zapastita movement in Mexico, when the movement demanded more say in the control of projects and reporting requirements and many NGOs pulled out as they could not balance this demand with upward accountability needs to donors (Andrews, 2014).

A 2014 UK study of investing in NGO monitoring, evaluation and learning systems raised similar concerns:

Our findings also showed that it is still common for analysis of project data to take place away from those who are implementing or benefiting from the projects, suggesting that accountability and communication to, and learning of, those further up the aid chain remains a higher priority than accountability and communication to, and learning of, beneficiaries and local organisations.

This dynamic is also reflected in capacity building where most intermediate and commissioning NGOs focus on ensuring that their implementing partners are able to collect the project data required for project management and accountability purposes, rather than considering the MEL needs of implementing partners more broadly (ITAD, 2014: 4)

The MEL needs of all partners need to be balanced. At the moment, the tendency is for the MEL needs of donors and implementing agencies including NGOs to take priority over those of local partners and governments.

The same UK study found that the full costs of MEL are not always budgeted for by the NGOs, partly because of concerns that the donors may not accept these costs. The NGOs studied also found it difficult to have sufficient capacity for monitoring and evaluation at a local, in-country level. Among the recommendations of the report were the need for NGOs to fully cost their MEL systems, ensure they are proportional to the value and use made of the information generated, and seek full cost-recovery. The recommendation for donors included the need to match their MEL funding to their expectations for MEL from the NGOs (ITAD, 2014).

In summary, monitoring and evaluation systems have developed rapidly over the last twenty years, but are not yet delivering a clear picture on what is working well and what is not working well in terms of development impact, because of the combination of issues discussed above. Some of these identified issues highlight weaker aspects of project level M&E systems which might be addressed by agencies. However, it leads us to consider that it may also be possible to answer questions around developmental outcomes and impact by considering different levels of thinking, questioning and learning beyond those that occur in single projects – that is, to monitor, evaluate and learn at and across different scales of work.

2.5 The challenge for NGOs: measuring at scale

Why demonstrate outcomes and impact at scale?

In addition to the broader need to explore this relatively uncharted territory of capturing, integrating and reporting on outcomes and impact across different scales, the NGOs also have particular needs and challenges that relate to their distinctive role in development. Communicating the significance, centrality and distinctive contribution of NGOs and civil society has been an ongoing challenge that needs a different and deliberate effort. The challenge for the 125 ACFID full members (2017) is how to articulate their particular contribution to a 'just, equitable and sustainable world', envisioned in ACFID's Strategic Plan 2015-20.

NGOs and civil society organisations (CSOs) may now be responsible for around a third of aid funds. Data is hard to come by and there is no centralised reporting system. Nevertheless, in 2009 it was estimated that NGOs were responsible (through their own funds and those they manage on behalf of official donors) for \$38.9 billion (32% of all ODA) (Riddell, 2013). The NGOs, historically considered peripheral to the global aid system, are significant in terms of their financial contributions. They can also be significant in terms of the alternative development processes they offer, as described below.

It is often forgotten that NGOs and CSOs predate the official aid system of post-war Europe and were the main providers of services to poor people in colonial times (Riddell, 2007). Their evolution since then has seen a great diversification in their roles, which still include service delivery, but also advocacy and holding governments to account, capacity development, community mobilisation and garnering public support for international development.

Some NGOs have been historically progressive in promoting alternative development paradigms over the mainstream linear theories of economic development. The diverse roles

of the NGOs and CSOs now numbering in their tens of thousands should be able to help inform the broader aid community on the effectiveness of complex aid mechanisms, particularly in addressing human rights, inequality and power relationships. It should be noted however that the degree to which NGOs are in fact engaged in transformative and alternative development paradigms has been in question over the past twenty years, with claims that ‘most NGO efforts remain palliative rather than transformative’ (Banks et al., 2015: 708). While no doubt much NGO work is ‘palliative’, it is accepted that alternatives to mainstream economic theories of development originate in the practice of NGOs.

The particular contribution of NGOs should not be measured in the same ways as all development partners:

‘NGOs are different to other aid delivery mechanisms....Critical for the relationship between NGOs and official donors such as AusAID is that recognising this difference can lead to making more effective use of NGOs in development and aid. But it also requires an understanding that NGOs have to be assessed from a more sophisticated perspective than many of the previous efforts have allowed. Donor mechanisms for performance review and assessment need to be varied to suit different aid delivery mechanisms. For NGOs in Australia, this means an understanding of how organisations combine their principles and program approaches with practice to lead to effective outcomes’ (Kelly and Chapman, 2010).

However, the perception that NGOs operate only at the margins persists with ‘the donor community’s narrow emphasis on NGOs and ‘results’” (Banks et al., 2015). The inability of the NGOs to articulate their empowerment strategies and their greater impact through government cooperation, national and international advocacy, networks and scaling up small approaches that work is also critiqued (Degnbol-Martinussen, 2003). Some consider the significance of the NGOs is not in their individual work empowering the poor or delivering services, but in the impact of their collective effect (Degnbol-Martinussen, 2003: 155), while M&E systems are often narrowly based and focussed at the project level. The SDGs are requiring collaborations beyond NGOs –with the private sector and governments for example – and this requires new ways of thinking about capturing results across collaborating partners and different scales of work.

Thinking ‘beyond the project’ is a major challenge for NGOs. Aid is organised primarily as projects and this poses serious challenges in measuring impact when there are hundreds of projects operating in a country at any one time. The fragmented and stand-alone nature of projects makes it difficult to assess what they ‘add up to’ (Degnbol-Martinussen, 2003). While evaluating projects may be able to capture individual efforts to strengthening capacity or other qualitative changes, it has so far failed to capture efforts by aid agencies to have broader impact beyond individual projects, or ‘aid’s strategic potential to influence local power structures, capacities and processes in developing countries’ (Degnbol-Martinussen, 2003: 266). Roger Riddell notes with some frustration ‘almost all the studies of the impact of NGO aid interventions continue to comprise assessments of individual, discrete projects’ (Riddell, 2013: 373). He goes on to suggest that some NGOs and CSOs in fact take a very narrow view of their work, confined to individual projects rather than ‘assessing how they might contribute to making a wider and long-term impact, where their absolute and comparative advantage lies, and how they might co-operate more with others’ (Riddell, 2013: 385).

In consultation with monitoring and evaluation practitioners from 11 Australian NGOs, and other stakeholders including consultants and representatives from DFAT and academia (ACFID, 4 May 2017), it was determined that there are multiple and complementary reasons for capturing, integrating and reporting on outcomes and impact at and across different scales beyond individual projects has the potential. These include:

- To help NGOs better communicate their individual or collective significance, centrality and distinctive contribution (as discussed above).
- To demonstrate that NGOs are legitimate development partners, on an equal but different footing to other development partners, and help demonstrate their strategic 'value-add'
- To help in influencing other development actors, and for advocacy, developing social movements and social change.
- For learning, within and between NGOs and other partners; to increase effectiveness.
- For accountability, to demonstrate and prove NGO work and help secure funding from institutional and public donors
- Efficient pooling of resources and being able to deal in longer time frames needed to demonstrate impact

What scales are we talking about?

The changing landscape of aid and development effectiveness as discussed in section 2.1, including the SDGs agenda, reinforces the need to work collaboratively with different types of actors, and potentially across different scales of work. This is consistent with the concepts of 'thinking and working politically', including brokering spaces for working more collaboratively. Such global-local partnerships may be operating at the levels of global advocacy and policy as well as at a village level.

Using the administrative rationale, where a project is the usual unit for the collection of activities contributing towards common development outcomes, and several projects form a program, different scales might include:

- across multiple projects within a single program delivered by the same agency
- across multiple programs delivered by the same agency
- across a single program delivered by multiple agencies
- across multiple programs delivered by multiple agencies

If we want to extend 'beyond the project' to include broader NGO and CSO contributions to impact, including within partnerships with other development actors, then it may make sense to start with a developmental rationale when considering the different scales. For example, climate change only makes sense when considered globally, people trafficking or avian influenza are clearly transboundary issues. Applying the developmental rationale, health, education, water and sanitation and others might be considered across sector, country or global levels.

A developmental rather than administrative rationale would take us away from thinking about single donors or single NGOs. However, there are deep tensions to be negotiated and challenges associated with this rationale. ANGOs have to negotiate their way through competing priorities within their organisations, including their international counterparts when relevant, as well as with donors and local partner organisations, communities and governments. There are conflicting demands, expectations and views about what is valuable to report on at or across scales as well as a continued debate about the priority of monitoring, evaluation and research compared to all the other things that need to get done. This negotiation inevitably involves different power dynamics. Helliker and Lissner argue that the negotiation tends to come out in favour of the NGO's own organisational stability and preservation (Helliker, 2007, Lissner, 1977). This supports the trend among the ANGOs consulted, who were mostly concerned with agency-level scales rather than country, sector or global scales (ACFID, 4 May 2017).

One concern identified by the ANGOs consulted was the strong incentive of upward accountability which detracts from locally connected and relevant learning. ANGO staff feel the imperative to achieve positive results and the pressure to work in locations with partners and clients that will bring positive results and value for money, compared with more marginalised areas and change that is harder to bring about. The risk of failure is seen as a risk to the agency's reputation, which is significant when it comes to competing over resources. The NGO environment is very competitive which can undermine collaboration and transparency across agencies. Finally, the exercise carries the risk articulated by the ANGOs 'what if it shows the wrong thing?' which reflects the focus on agency reputation as opposed to learning purposes.

Capturing, integrating and reporting at and across different scales beyond an individual organisation requires high level strategic intent by the organisation, with associated commitment, processes and capabilities. One inherent issue is that monitoring, evaluation and learning roles in the ANGOs are often narrowly focussed on technical skills. This can make negotiating organisational buy-in to a different approach difficult. In some ways, starting with monitoring, evaluation and learning across different scales is premature because it makes most sense to do this when agencies are *programming* at and across different scales. However, there is a chicken-and-egg element to this – if the benefits of capturing, integrating and reporting across different scale are visible and obvious, it can help drive programming behaviours.

2.6 Factors to consider when capturing, integrating and reporting on outcomes and impact across different scales

This chapter has started to address the breadth of the political and practical challenges in capturing, integrating and reporting on outcomes and impact across different scales. Situated in a shifting context for development cooperation and different understandings of development effectiveness, project level monitoring and evaluation systems have not adapted to the needs of addressing complexity and politics and are largely not delivering on deepening the understanding of what is/is not working well in terms of development impact. The combination of an overly technocratic approach, project fragmentation, the need to address complexity and politics, and to balance partner needs and ensure the costs and benefits are commensurate and covered have let down current M&E systems.

It may also be possible to answer questions around developmental outcomes and impact by considering different levels of thinking, questioning and learning beyond those that occur in single projects – that is, to monitor, evaluate and learn at and across different scales of work. Expanding thinking 'beyond the project' includes thinking beyond administrative boundaries. To the extent possible we should use a developmental rather than administrative rationale to determine the scales for capturing, integrating and reporting across different scales. If we continue to think in project mode, there is a risk that monitoring and evaluating at different scales will simply exacerbate the pitfalls of M&E systems. There must be a very clear purpose in mind before embarking on what is likely to be a technically quite complex exercise which will require additional resources to implement.

A number of risks are inherent:

- There is a risk that we might end up promoting even more technocracy in an already technocratic arena.
- There is a risk we can make this whole exercise extremely complicated and exacerbate the issues of fragmentation and crowdedness even further.
- There is a risk that we oversimplify the complexity of development processes and what the NGOs bring to this, and that we discuss the NGO contribution in the same terms of any other actor.

- Depending on the nature of the scaled-up reporting we are talking about, possibly the worst of the risks is that we embark on this undertaking to the exclusion of the partners and beneficiaries and develop the capacity among the NGOs in Australia without developing the capacity of those at the front line who also need to make sense of their collective contributions.

By way of a summary of the findings of this chapter, the following list of factors are proposed for consideration when contemplating the design of a process to capture, integrate and report on outcomes and impact across different scales.

- Being clear on the purpose and questions and making use of the information
- Applying fit-for-purpose methods in fit-for purpose ways
- Adopting methods and approaches that address complexity
- Considering the needs of all partners, including locally
- Capturing the distinctive contribution of the NGOs
- Not being overly complicated, technocratic or exacerbating fragmentation
- Having adequate resourcing to meet the purpose

These factors are applied in the next chapter, which describes and discusses some of the methods and approaches available in the application of this process.

3 Methods and approaches for capturing, integrating and reporting on outcomes and impact across different scales

Being able to capture, integrate and report on evidence of outcomes and impact across different scales is going to require different approaches, methods and mindsets than 'traditional' project M&E. As detailed in Chapter 2 above, the scales addressed in this report refer to both administrative and developmental framing, which will each inform the appropriate method or methods to apply.

This chapter looks at the strengths and limitations of the various approaches and methods, taking account of the seven factors to consider when capturing, integrating and reporting on outcomes and impact across different scales:

- Being clear on the purpose and questions and making use of the information
- Applying fit-for-purpose methods in fit-for purpose ways
- Adopting methods and approaches that address complexity
- Considering the needs of all partners, including locally
- Capturing the distinctive contribution of the NGOs
- Not being overly complicated, technocratic or exacerbating fragmentation
- Having adequate resourcing to meet the purpose

The five case studies examined for this research (see Annex 3) all addressed different scales and had different, but clear, purposes as shown in Table 1. The strengths and limitations of the methods they adopted are discussed under the relevant methods section below.

TABLE 1: SCALES, PURPOSES AND APPROACHES OF THE FIVE CASE STUDIES

Case Study	Scale	Purpose	Approach
1. Tracking Impact: An exploratory	Single donor (NORAD) across all	To assess the overall and wider impact of	Strategic evaluation

study of the wider effects of Norwegian civil society support to countries in the South	supported agencies, projects and programs in 3 countries	NGOs and CSOs on development and their long-term contribution to poverty reduction	
2. Reach, Relationships and Results: Case Studies of Australian NGOs' work in Education	Single sector (education) across multiple projects and programs, countries and agencies	To demonstrate to DFAT the unique and necessary contribution of NGOs to education	Research approach: Case studies
3. Exploring the link between child and youth participation and development effectiveness	Single development approach (child and youth participation) across three projects, three agencies and in-country partners in three countries	To investigate the link between child and youth participation and development effectiveness	Research approach: Qualitative methods
4. Australia Africa Community Engagement Scheme (AACES)	Single DFAT funding window in three sectors across eleven countries, ten agencies and in-country partners	To meet DFAT's accountability requirements and strengthen the program to target and serve the needs of marginalised people	Combined approach: donor-led
5. OXFAM Australia's approach to reporting against its Strategic Plan (2014-19)	Single agency, all projects and programs	To monitor the overall performance of OXFAM Australia programs; contribute to thematic strategy, program learning and program development; and support engagement with institutional donors, supporters and the public.	Combined approach: agency-led

Many of the methods and approaches used to capture, integrate and report on outcomes and impact across different scales overlap with those used for project level M&E and are the topic of text books, training courses and websites (such as Better Evaluation). This chapter is not intended to be a comprehensive treatise on the topic, but rather a high level overview of selected methods, illustrated by the case studies and by other examples that have come to light in the course of the research. The selected methods are chosen to reflect the breadth of available approaches with practical examples. They demonstrate how the seven factors apply in identifying strengths and limitations.

Three main groups of methods (indicators, evaluation and research) are described and analysed as a means to define a categorisation of broad approaches. However, it is acknowledged that the boundaries between the three are blurred and overlapping.

The first group of methods is about using indicators, including aggregating results and some of its possible variations. Indicators are often the basis of monitoring and evaluation systems so there is overlap with the second group of evaluative methods. This group of methods includes both strategic level evaluations and evaluation synthesis. Evaluation is a sub-group of research more broadly and there is therefore considerable overlap with the third group of methods which are more broadly research-related. The difference is in the intent – more judgmental in evaluation and more investigative in research. These three approaches are often applied in combination. A deeper look at how this is done in the last two case studies is discussed.

3.1 Using indicators

There is a long history of using indicators in development cooperation, a practice increased by the Paris ‘results’ agenda (Holzapfel, 2014). The use of common or standard indicators in NGO monitoring, evaluation and learning systems is a popular approach. In the UK for example, BOND (the UK membership body for non-governmental organisations working in international development) has developed the *Impact Builder*, an online sector wide framework of outcome indicators, data collection tools and assessment methods to improve the consistency of how UK NGOs measure, learn from and communicate results as a stream of work in the BOND Effectiveness Programme (BOND). The strengths and limitations of using indicators in M&E systems are described elsewhere (for example (Holzapfel, 2014)), this section looks specifically at their use across different scales of work. It first describes a number of different ways that using indicators can be applied across different scales and then looks at the strengths and limitations of the various approaches.

3.1.1 Aggregating results

Adding up results from standard and clearly-defined indicators across projects and across countries is a popular approach, particularly in terms of measuring beneficiary numbers, which can help demonstrate the reach of an NGO and the scale of its work. The approach requires that the same indicator definition is used.

World Vision International for example has developed a ‘compendium of indicators’ of child well-being outcomes, drawn from both globally-accepted standard indicators as well as World Vision ‘innovative’ indicators. World Vision project managers can select from this compendium according to what is most relevant for the project and appropriate for the local context. Over 200 indicators, covering both quantitative and qualitative measures are included as well as some particular tools for measuring (for example survey guidance). The compendium allows project managers to change the indicator to match the one used by the government of the country as ‘World Vision should seek to support existing data collection efforts, rather than create a parallel measurement system’ (World Vision, 2014).

3.1.2 Framing indicators

A less centrally driven approach which does not require projects to use standard indicators is the use of ‘framing’ or ‘basket’ indicators’ where similar and related project level indicators are collected under a single global indicator and summarised according to a global theme. An example given in a useful INTRAC paper shows how project indicators B, C and D might be collected and described under a global indicator A:

Global indicators	Project indicators
CSOs influence government practices at local levels (A)	# of times governments invite CSOs to attend meetings to discuss policy (B) Evidence of CSO submissions being copied into government policy document (C)

Successful lobbying of local government by CSO
(D)

Source: (Simister, 2016)

3.1.3 Coalition indicators

Indicators are commonly used for measuring the effectiveness of coalitions, networks and partnerships, often in two ways – the quality of the partnership or coalition itself as well as the results of the collaboration. The BOND *Impact Builder* described earlier provides some examples of standard indicators that can be used to measure the strength and effectiveness of advocacy coalitions (Table 2).

TABLE 2: EXAMPLE OF INDICATORS OF COALITIONS IN ADVOCACY (BOND)

Outcome: Coalitions/networks are stronger and more effective	
Sub Outcomes	Indicators
Coalition/networks have improved capacity to coordinate and influence	Network / coalition demonstrating improved capacity and practice in coordinating and undertaking collective advocacy
	Overall improvement in the effectiveness of the network/coalition
	Improvement in a certain aspect of the network/coalition
Coalition / network has influence	# and frequency of instances where power holders proactively approach coalition for inputs / information
	# and frequency of instances where power holders proactively approach coalition to request meeting
	# and type of media organisation proactively contacting (organisation x) for comment or advice on (issue y)
Increasingly organisations are working collaboratively	# organisations working on (issue x)
	# of other organisations' work plans, priorities, policy positions and statements influenced by organisation's position on (issue x)
	Instances of joint working between organisations on (issue x)

3.1.4 Rating indicators

When standard ratings are used at project level, these can be analysed at a higher order to understand more about the quality or other characteristics of a whole project portfolio. The CARE 2020 Program Strategy indicates that all programs should promote gender equality and women's voice. CARE tracks the integration of gender into its projects and initiatives through a Gender Marker. Each project answers a set of common questions which result in a ranking of whether the project is gender transformative, gender responsive, gender sensitive, gender neutral or gender harmful. CARE can analyse the Gender Markers globally to track progress against implementation of its Program Strategy and determine, for example, where more support in integrating gender might be needed.

In another example, DFAT project managers rate their activities using the guidance in a standard matrix in the Aid Quality Reporting system. The Office of Development Effectiveness analyses these ratings across the whole agency. An example from this analysis compared ratings for effectiveness and sustainability for different sectors across

DFAT. The analysis showed, for example, that sustainability and effectiveness in ‘improved government’ projects were poor compared to other sectors (Office of Development Effectiveness, 2014b). Such a finding can be used to focus attention on changing approaches and improving performance in this sector.

3.1.5 Strengths and limitations of using indicators

Being clear on the purpose and questions and making use of the information

The indicator-based methods described above are often used in combination and have wide appeal in aggregating beneficiary numbers among the international NGOs, some of whom have developed quite sophisticated infographics and country-specific case studies on their websites to demonstrate results, combining quantitative and qualitative data. A fairly typical example of the kind of information generated is for Christian Aid in the UK, for Burundi:

In the last year, 70% of 21,121 people living with HIV surveyed indicated being members of a support group (up from 46% at baseline), and reported benefits including increased income, access to information, moral support and reduced stigma and self-stigma (Christian Aid).

Indicators on their own do not tend to measure particularly meaningful things however – even with the example above, we do not know if the 21,121 people surveyed is significant in terms of those living with HIV in Burundi, whether the support groups are locally managed and sustainable, what exactly Christian Aid did to bring about this result or how this activity interacts with other initiatives in Burundi to bring about lasting reductions in stigma.

Aggregating results applies to the literal adding up of results and is most useful for service delivery purposes, and easier at output level than outcome level. The actual use of the information generated may be limited to donor reporting, but also for Annual Report publications, websites and public relations materials rather than for learning. There is a tested benefit of using these types of indicators in public relations materials for fundraising although whether they actually satisfy the needs of aid supporters may be in question. Horton, for example, asks:

I am not an aid specialist, but a moral philosopher. Like many other moral philosophers, I believe that a strong case can be made for thinking that those of us who are relatively well-off, in global terms, should give some of our money to aid agencies such as Oxfam, World Vision, and CARE. Certainly, if the effects of the work aid agencies do are anything like as positive as their fundraising literature implies, then the case for thinking that we should give would be very strong indeed. Are the effects as positive as that, though? (Horton, 2010: 27)

To some degree the use of framing indicators allows for a more qualitative discussion of NGO contribution if the information is used as part of a broader performance story (Mayne, 2004). Rating indicators can help to capture some of the particular contribution of NGOs and be used to help manage program portfolios.

However, used on their own, indicators are insufficient in capturing, integrating and reporting across different scales. It is not necessarily enough to tell the public ‘that probably as many as 80% of NGO aid projects achieve their immediate objectives – clinics were built, teachers were trained, children were immunised and water projects provided clean water to those who needed it – what they increasingly want to know today is what difference these individual successes made to overall development and what contribution NGOs make to sustainable development and the reduction of poverty at the country level.’ (Riddell, 2013: 374).

Applying fit-for-purpose methods in fit-for purpose ways

The use of indicators beyond discrete projects or areas in development cooperation is subject to limitations and these apply (and are sometimes amplified) when using indicators at and across different scales:

.... indicators cannot be applied equally well to all intervention contexts and their use may lead to distortions. The use of indicators is particularly likely to have adverse effects where they are used to assess performance. For instance, there is the risk of resources being shifted to areas where performance is more easily measurable or where results are easier to achieve. Moreover, accountability demands may lead to too much emphasis being placed on results measured by indicators at the expense of unquantified performance aspects. (Holzapfel, 2014: 154-5)

It is possible and desirable to collect both quantitative and qualitative indicator information (including converting qualitative information into quantitative), but there is a tendency for the NGOs to focus on quantitative information. A risk identified in the UK study of NGO monitoring, evaluation and learning systems is that quantitative data is easier to store and analyse than qualitative data:

Two of the five top level case-study NGOs mentioned that they find qualitative data collection more of a challenge because their MEL systems were more tailored towards quantitative data, as this is easier to collect and sometimes preferred in donor reporting (ITAD, 2014)

There is a risk in using indicators that the means of collecting the data drives the method, rather than the purpose.

Adopting methods and approaches that address complexity

Indicator-based methods on their own do not address complexity. One obvious issue with adding up results across different project contexts is the level and type of effort required to achieve a particular result (say immunisation of a child) is quite different in a remote area of Papua New Guinea than in Port Moresby, which in turn is quite different again than for a child in Hanoi. Aggregating indicators across different scales can be misleading and oversimplify development efforts.

Using indicators to measure partnerships can also oversimplify impacts, constraints and benefits because partnerships are constantly evolving processes. Similarly the indicators can be interpreted differently by the different partners and while they might provide a forum for dialogue, they might also provide a source of conflict (Caplan and Jones, 2002).

Considering the needs of all partners, including locally

There are assumptions that this type of reporting is demanded by the donors and indeed for several years, the Australian Government has required reporting of specific aggregated results. Case Study 4 (The Australia Africa Community Engagement Scheme) specifically included a mechanism to collect and aggregate such a set of results in response to this felt donor need.

Case Study 5 (OXFAM Australia's Reporting Approach) includes an element of using indicators to meet donor requirements and to manage their project portfolio. Collecting indicator information can help ANGOs get a clear picture of the scale of their particular portfolio and particular characteristics about the range of projects which can help in portfolio management.

However, the data usually sits with the INGO for analysis and publication, such that local ownership and alignment with local systems is not usually a consideration. Local ownership of indicators using participatory M&E approaches can support more empowering than 'extractive' processes (Mayoux and Chambers, 2005).

Capturing the distinctive contribution of the NGOs

While potentially useful in illustrating something quite tangible, aggregating results is not actually saying anything about the contribution of the NGO to achieving that result, nor its sustainability, changes in power dynamics, performance or contributions to development more broadly. If NGOs see reaching the maximum number of beneficiaries as reflective of their impact (it is, for example, among the criteria in a global ranking of NGOs by a Swiss-based media organisation (NGO Advisor)) then it can be useful for them. However, interpreting NGO effectiveness in terms of directly lifting every person in the world out of poverty is inconsistent with promoting sustainable change which addresses the causes as well as the symptoms of poverty and marginalisation – i.e. reduces poverty and builds capacity within communities, civil society and government to address their own development priorities. Again, participatory methods in which participants are actively shaping and using information are more aligned with NGO effectiveness principles, but such methods are not the norm (Mayoux and Chambers, 2005).

Not being overly complicated, technocratic or exacerbating fragmentation

One strength of this approach of aggregating results is the ease and simplicity of presenting results aggregated from many interventions. Another strength is in its promotion of standard, universal indicators and when these are aligned with those used in the country where the projects operate, it can help guard against a massive proliferation of indicators, at least in theory. In practice, there has been a tendency to develop large numbers of indicators in project-level M&E frameworks, and because each project has its own unique framework, a proliferation of indicator-based M&E frameworks now exists at the country level.

Having adequate resourcing for the purpose

Collecting indicator information requires standard application which means investing in training for staff to ensure the precise definitions are understood and applied. Other investments are in software to manage the information systems – while many NGOs use existing software (such as Excel), others have invested in bespoke systems. The costs of investing heavily in this type of performance measurement should be balanced with the benefits and the need to invest in other types of performance measurement.

Quantifying outcomes and impact in both quantitative and qualitative terms is very important. But the popularity of an agency-centric use of aggregating indicators or related approaches among international NGOs is cause for some unease, because on their own indicators do not address complexity, are usually not considering the needs of local partners and are not helping to further a global understanding of the distinctive contribution of NGOs. One issue is the extractive rather than empowering use of indicators. Another issue is the isolated use of indicators in reporting. They are better used as part of telling a ‘performance story’, a reporting method developed by John Mayne (Mayne, 2004). The main story line of a performance story is how well a project or program has performed in relation to what was expected and what will now be done differently to better ensure future performance. This idea of performance stories was designed with single interventions in mind, but it contains an important principle of not only reporting the outcomes but also discussing the evidence available to demonstrate the contribution made by the intervention to those outcomes. This principle can also be applied at an aggregate level if the use of indicators is integrated into a narrative, or into evaluative or research approaches, discussed next.

3.2 Evaluative approaches

Evaluation is a type of applied research that usually involves a judgement of value, merit or worth. The Australian Evaluation Society uses the following definition:

While many definitions of evaluation are used, the term generally encompasses the systematic collection and analysis of information to make judgements, usually about

the effectiveness, efficiency and/or appropriateness of an activity (Australian Evaluation Society, 2013: 3).

Evaluations are commonly undertaken by development agencies for the ‘assessment, as systematic and objective as possible, of an on-going or completed project, programme or policy, its design, implementation and results’ (Development Assistance Committee, 1991). Evaluation methods have evolved rapidly and evaluation is increasingly viewed as an academic discipline in its own right (see for example (Scriven, 1994, Pawson, 2014)). The Better Evaluation website outlines twenty-two different approaches to evaluation, from appreciative inquiry through outcome harvesting to utilisation-focused evaluation (Better Evaluation). Theory-based evaluations are based on testing theories of change and their underlying assumptions and generally will seek to understand more about why something is or is not working. They are particularly relevant for the work of civil society and the learning objectives of evaluation.

There are basically two ways that evaluative methods can be applied to help capture, integrate and report on outcomes and impacts across different scales of work. The first is to undertake an evaluation that targets the particular scale or scales rather than a single project. We are calling this a strategic evaluation. The second is to synthesise information available in existing project-level evaluations. These are discussed in turn.

3.2.1 Strategic evaluation

A higher order evaluation can be used to evaluate programs, approaches, agencies, multiple projects or to evaluate a particular theme. The main strength of a strategic evaluation approach is that it applies the rigour of evaluative methods at a higher order, usually to generate lessons in implementation and evidence of impact. It is likely to draw on existing monitoring or evaluative information as well as undertaking some primary research to answer the specific evaluation questions. A strategic evaluation approach can be at or across any level or scale. DFAT’s Office of Development Effectiveness has undertaken many strategic evaluations. One example, across a thematic area is its evaluation of *Violence Against Women in Melanesia and East Timor* (2008).

Another thematic example is the World Bank’s 2016 *Evaluating the Evidence for historical interventions having reduced HIV incidence: a retrospective programmatic mapping modelling analysis* (The World Bank, 2016). It focuses on understanding the relative effectiveness in quantitative terms of antiretroviral therapy (ART) and changes in sexual risk behaviour across five countries.

Case Study 1 *Tracking Impact: An exploratory study of the wider effects of Norwegian civil society support to countries in the South* (Annex 3) is another example of a strategic evaluation⁵ – in this instance to assess the overall and wider impact of NGOs and CSOs on development and their long-term contribution to poverty reduction across four countries.

The analysis of strengths and limitations in applying a strategic evaluation approach draws on these and other examples, but largely focuses on Case Study 1.

3.2.2 Strengths and limitations of strategic evaluation

Being clear on the purpose and questions and making use of the information

⁵ This is considered in this paper as a strategic evaluation because the purpose was judgmental (although it describes itself as an ‘experimental approach’).

Having a clear purpose helps to define the evaluation questions which then determine the particular method. The purpose of Case Study 1 *Tracking Impact* was to assess the overall and wider impact of NGOs and CSOs on development and their long-term contribution to poverty reduction. It investigated the performance story and particular contribution of NGOs at a high level in four countries – Ethiopia, Malawi, Nepal and Vietnam. The study itself is published (Abuom et al., 2012) as well as an analysis of the approach used, by one of the Panel members (Riddell, 2013). However, how the evaluation was used by NORAD or other stakeholders is not described.

Applying fit-for-purpose methods in fit-for purpose ways

The evaluative method should match the purpose and objectives of the evaluation. Here we should mention the range of evaluation paradigms that exist in this rapidly growing field. At one end of the spectrum (and currently the dominant view) are the firm believers that impact cannot be proved unless there is a counterfactual (i.e. what would have happened without the intervention). These evaluators prefer the use of experimental or quasi-experimental methods involving randomised control trials. At the other end are the realists who recognise the messy complexity of trying to understand how and why change happens and adopt corresponding social science methods to unravel and understand this complexity. Between these two extremes are many other useful paradigms, evaluation approaches and methods. There is a 'Choosing Appropriate Evaluation Methods Tool' available on the BOND website⁶ for selecting evaluative methods according to the types of questions being asked (although it is not clear if the quality of this tool has been tested.)

The Panel in Case Study 1 had a clear purpose and grouped what they called 'development results' in three categories: Improved quality of life for individuals and communities; wider and long term effects; and strengthened organisational capacity. They developed 19 working hypotheses and 56 questions to guide the assessment. Because the first category is often the subject of other work, the focus was more on the second and third categories.

Each country visit lasted one week. It involved meetings and interviews with Norwegian CSOs and their partners, with government representatives, independent researchers and journalists, and with other International NGOs and other donors. The country work also involved reviewing documents, reports, studies and evaluations, especially those shedding light on the wider and long-term impact questions. Workshops with key local CSOs also took place, where the relevance and validity of the proposed hypotheses were discussed.

Among the methodological changes the Panel proposed if such a study were to be repeated, were to undertake a longer and more systematic analysis of the project documentation and a more in-depth and wide-ranging review of the relevant literature.

Adopting methods and approaches that address complexity

This factor is one of the most significant in selecting among evaluative methods. A number of evaluative methods embrace complexity. An interesting example is an ethnographic evaluation across three programs in NGO-led HIV prevention and sexual health programmes with rural youth in Uganda (Bell and Aggleton, 2012). This offers a practical example of using rigorous social science methods which embrace the realities and complexities of the interventions and the settings in which they operate. Contribution analysis is a promising method for addressing both complexity and attribution issues (Mayne, 2001) and can be used in combination with other evaluative methods. Realist evaluation (Pawson and Tilley, 1997) also embraces complexity and has been proposed in a Finnish study as an avenue to

⁶ <https://www.bond.org.uk/resources/evaluation-methods-tool>

learning for development NGOs (Holma and Kontinen, 2011). Case Study 1 Tracking Impact also embraces complexity to some degree.

It is useful here to consider the findings of the World Bank's *Evaluating the Evidence for historical interventions having reduced HIV incidence* (which does not address complexity) to appreciate the significance of evaluation method selection. The World Bank case was a multi-country study focused on evaluating whether ART scale-up and changes in sexual risk behaviour have contributed to the declining trends of HIV incidence and prevalence. Using mathematical modelling, it found differing results in the five countries (Botswana, Dominican Republic, Kenya, Malawi and Zambia) engaged in the study. In Botswana for example,

'...there was strong evidence that showed that ART and changes in sexual risk behaviour had an impact of averting 210,000 infections in urban areas and 120,000 infections in rural areas (1975-2012). This discrepancy between urban and rural area results was thought to be due to geographical heterogeneity in HIV epidemiology or lack of power in available data' (The World Bank, 2016)

This particular method speculates but does not specifically investigate the reasons for differences between urban and rural areas in Botswana nor the finding that ART was having a marginal impact on prevalence trends. A large study such as this is likely to have influence on future program decisions, but these will not be informed by an understanding about why particular interventions work better in some contexts. Methods embracing complexity would be more likely to investigate the reasons behind the findings.

In contrast, the Panel in Case Study 1 remained very cognisant of complexity and, in particular, the different contextual settings. One of their recommendations if such an exercise were to be repeated was to have longer country visits in order to have 'more in-depth discussions with different stakeholders and with more beneficiaries, and to be able to scrutinise more carefully the differing and at times conflicting views on the impact of CSOs provided by different informed interlocutors and stakeholders; to pursue more vigorously the central issue of attribution; and to provide more documentation of why particular conclusions were drawn and judgements reached, especially where data and interviews produced differing assessments' (Riddell, 2013: 388). The proposed improvement to the method was a recognition of the complexity of the different contexts.

Considering the needs of all partners, including locally

It is not entirely clear how the particular needs of local CSOs were considered in Case Study 1. However, based on the Panel's firm view that the wider impact of CSOs can only be assessed from the basis of an in-depth understanding of the context, culture and history of each country, they recommended that experienced evaluators and scholars from each country are included in any future Panel of assessors. (Riddell, 2013: 388). There was no discussion about the needs of the local CSOs for information about their own collective effectiveness.

Capturing the distinctive contribution of the NGOs

Case Study 1 was specifically designed to capture the contribution of CSOs, and in particular their impact 'beyond the project'. The (revised) approach holds considerable promise for assessing the wider effects of NGOs and CSOs on development at a country level.

Not being overly complicated, technocratic or exacerbating fragmentation

The method for Case Study 1 cut through the technocracy of monitoring and evaluation to focus squarely on the question of what it all 'adds up to' in a meaningful and potentially rigorous way.

Having adequate resourcing for the purpose

Strategic level evaluations tend to be expensive and complex undertakings with methodological challenges. They usually comprise a specialised evaluation team over a time period of at least a year. Case Study 1 covered four countries in a very short time frame, but recommended much longer time was needed in future applications. The cost is not known.

Using a strategic evaluation to evaluate across different scales is a complex exercise that can be methodologically challenging and expensive. However, if the purpose and evaluative questions warrant such an exercise and align with methods that address complexity (for example ethnographic evaluation, contribution analysis, realistic evaluation, or an adapted panel approach) it can provide explanatory insight into outcomes and impact across higher orders of scale than individual projects. A strategic evaluation, as in Case Study 1, will often draw on existing documentation and evaluations as well as complement this with primary research. In this sense evaluation syntheses, discussed in the next section, are not always applied as standalone methods but may be integrated into strategic evaluation designs.

3.2.3 Evaluation synthesis

There are a variety of methods available to synthesise information across existing project evaluations. We should begin with a note however about terminology. This section could be called 'systematic review' as it refers to the comprehensive, systematic and transparent synthesis of existing evidence (Snilstveit et al., 2012), but the expression 'systematic review' has taken on a shorthand meaning in some evaluation literature to mean the systematic review of randomised control trials. Also, because we are looking specifically here at syntheses of existing evaluations as opposed to other types of research, the title 'Evaluation syntheses' is chosen. But these systematic review methods can equally be applied to non-evaluative evidence as research approaches.

Better Evaluation outlines nine different technical approaches (Better Evaluation) to synthesising project evaluations. These can include quantitative approaches, such as adding up the number of positive and negative evaluations (vote counting), or meta-analysis, a statistical method for combining numeric evidence from experimental or quasi-experimental studies to produce a weighted average effect size. These quantitative methods can be used to demonstrate impact across different scales, but do not explain the findings if used on their own and also run the same risks as aggregating indicators in terms of adding things up across different development contexts. Depending on the purpose of the exercise, there will often be benefit in using mixed methods. Including qualitative approaches can help in explaining the effects, and this is increasingly being used to add more explanatory power to quantitative methods (Snilstveit et al., 2012). Qualitative methods can also be used to generate quantitative data (Mayoux and Chambers, 2005).

This section will first describe and then consider the strengths and limitations of three qualitative methods: systematic review, realist synthesis and variations of textual narrative synthesis, as all have relevant examples from aid and development and they reflect the range of philosophical approaches underpinning the methods. A more comprehensive discussion of the available approaches and methods is available (Snilstveit et al., 2012).

3.2.3.1 Systematic review of randomised control trials

Trends in impact evaluation in international development have favoured experimental designs in an attempt to prove causality and attribution using counterfactual data. Such use of randomised control trials currently dominates the social sciences. Several donors combined their resources to establish the International Initiative for Impact Evaluation (3ie) in 2008, probably the biggest such grant-making body, that supports impact evaluations that meet its rigorous standards. Of particular interest here is the work 3ie has done in its systematic reviews, using rigorous and transparent methods to synthesise studies (usually

randomised control trials) of the same topic to answer specific questions. The main question is usually around ‘what works?’

The largest of these reviews to date, in terms of the numbers of included studies, looks at education interventions in low and middle income countries. With a focus on randomised control trials, the initial search of 79,000 documents was culled to 420 papers covering 216 programs in 52 countries. The synthesis also drew on 121 qualitative studies but this did not alter the methodological reliance on experimental data (Snilstveit et al., 2016). The main finding from the systematic review of education interventions can be summarised:

Some interventions, such as cash transfers and structured pedagogy, appear to work in most contexts, while the effects of other interventions, including different types of school-based health programmes and reducing school fees are yet unknown due to lack of sufficient studies (Snilstveit et al., 2016: 51).

3.2.3.2 Realist syntheses

In stark contrast, realist syntheses make use of all available evaluations providing the particular piece of information is supported by sound evidence. Pioneered by Pawson and Tilley (Pawson, 2006, Pawson and Tilley, 1997) it is a relatively recent method that seeks to uncover the mechanisms that make projects work in some contexts but not others. The focus of a realist review (as for a realist evaluation) is in answering the question ‘what works for whom, under what circumstances and why?’

A recent example asked the question ‘How do community-based conservation programs in developing countries change human behaviour? This realist synthesis looked at 17 conservation project evaluations and found three mechanisms operating, depending on contextual factors:

- i) conservation livelihood provides economic value;
- ii) conservation provides benefits that outweigh losses of curtailing previous behaviour; and
- iii) giving local authority over resources creates empowerment (Nilsson et al., 2016).

It identified the particular aspects of the context (such as the cultural acceptability of the type of livelihood) that affected the degree to which each of these mechanisms succeeded. It also found that the commonly used activity of education in conservation did not change behaviour.

3.2.3.3 Textual narrative syntheses

A variety of methods can be applied to a narrative synthesis of evaluative texts. These can include content analysis (for example frequency analysis of common themes), framework synthesis (which often starts with a concept framework for analysis and refines this as the data emerges), ethnographic approaches and others (see (Snilstveit et al., 2012) for a fuller description). Undertaking narrative syntheses of evaluation texts can make use of software to help with analysis of both quantitative and qualitative data. They can be less technically demanding than the previous two methods (noting that realist synthesis can be considered a form of textual narrative synthesis), but have been subject to critique for not being sufficiently transparent and care should be taken to explain the exact method used. Narrative syntheses that are applied to non-evaluative materials can be equally applied as research methods (considered in the next section).

Often, some type of meta-evaluation which evaluates the quality of the evaluations to be synthesised is first undertaken. For example, in 2014 ODE synthesised project evaluations from across the agency in its report *Learning from Australian Aid Operational Evaluations* (Office of Development Effectiveness, 2014a). Following a meta-evaluation assessing the

quality of DFAT's operational evaluations, the 64 evaluations that had credible evidence were analysed for 'what works well and why and what doesn't work well and why'. The most strongly emerging common findings are reported. This example is for all evaluations across the whole agency.

Similar products are prepared by some NGOs, for example World Vision Australia's *Learning From Experience: Annual Evaluation Review 2015* or OXFAM Australia's *Evaluation Reviews*. A number of other examples illustrate how the approach can be applied in its simplest form, to synthesise findings across similar approaches within the same agency:

- *Bringing Innovation to Scale: Resilience to Climate Change: Synthesis of learning from four CARE Community-based adaptation projects*. A synthesis of findings from 4 project evaluations with lessons for future programming
- *World Vision Channels of Hope Solomon Islands 2016*. Narrative synthesis of findings from evaluations of gender norms change and violence prevention initiatives in Honiara, Temotu and Weather Coast

3.2.4 Strengths and limitations of evaluation synthesis

Being clear on the purpose and questions and making use of the information

A realist synthesis can generate findings that are useful for informing how initiatives might be designed and managed in different contexts, as the method embraces complexity and the recognition that different contexts allow different mechanisms to work is its fundamental premise. If the purpose of the evaluation synthesis is to find out what works well and why and there are existing evaluations to synthesise, this method would be appropriate.

A systematic review of randomised control trials is more geared to answering the question of what works well in every circumstance. Again, if this is the question being asked, and at an appropriate scale or scales, the method is appropriate. However, one of the weaknesses of this approach is in the limited usefulness of its findings, because when applied at a global scale it is seeking to find out what works in every context.

Variations of a textual narrative synthesis can be used to answer particular questions and purposes, provided the underlying data being synthesised contains the required information. Often project evaluations contain insufficient information particularly around the links between the intervention and the outcomes.

Applying fit-for-purpose methods in fit-for purpose ways

The three groups of methods described here range from complex to relatively straightforward, providing a range of options to suit a particular evaluation purpose and questions.

Adopting methods and approaches that address complexity

Realist syntheses by definition embrace complexity. Systematic reviews of randomised control trials have been criticised as methodologically inappropriate for most social programs, for their screening out of context as well as lack of concern for investigating the reasons that interventions work (Ramalingam, 2013, Pawson and Tilley, 1997). The degree to which textual narrative syntheses address complexity likely depends on the particular method.

Considering the needs of all partners, including locally

None of the examples reviewed considered the needs of local partners.

Capturing the distinctive contribution of the NGOs

NGOs have not tended to use randomised control trials, so systematic reviews of randomised control trials are less likely to be relevant for their needs. Depending on the purpose, questions and underlying evaluations, both realist syntheses and textual narrative synthesis could be used to capture the distinctive contribution of NGOs.

Not being overly complicated, technocratic or exacerbating fragmentation

These methods use existing evaluations so will not exacerbate fragmentation. They need not be overly complicated although they all require high levels of expertise.

Having adequate resourcing for the purpose

Textual narrative syntheses depend on the quality of the evaluations being synthesised. If the source material is inconclusive then the synthesis will not be able to draw conclusive findings. Depending on the particular questions being asked of the synthesis, they can focus on the particular contribution of the development approach or NGO. Findings can inform future programming or be used to demonstrate the effectiveness of a particular approach. They need not be expensive exercises and can potentially help to make sense of a number of evaluations.

Systematic reviews of randomised control trials require a particular quality in the evaluations being synthesised as illustrated with the example. Realist syntheses try and make use of all available information, provided there is sufficient evidence for the particular piece of information being used. Nonetheless, both types of syntheses are likely to require specialised team members. The expense this entails might be overcome with pooled resources. In any case, as they are drawing on existing evaluations, there is less cost involved in gathering primary data.

As already described, the use of various evaluation synthesis methods can complement strategic evaluations or can be standalone. Realist synthesis and other textual narrative syntheses potentially address complexity and can make good use of existing evaluative information. The exercises require careful and transparent methods. Because they draw on existing evaluations there is likely a time and cost saving compared to a strategic evaluation. However they are also dependent on the quality of the underlying information.

3.3 Research approaches

The Research for Development Impact (RDI) Network uses the Australian Code for Responsible Conduct of Research 2007 definition of research: 'an original investigation undertaken to gain knowledge, understanding and insight' (ACFID, 2016: 2). The methods already described under evaluation are research methods, and the methods described here under research might equally be applied to evaluations, but the intent and purpose is what distinguishes research from evaluation. If the intent is more judgmental, involving assessing, then an evaluative approach is more useful, if the intent is more to gain knowledge, understanding and insight then a research approach is appropriate.

Research methods and tools may draw on different social science, anthropological, political science, economic or other research methods. It is an endless field of research methods that may be used to analyse or synthesise information across different scales without the source material adopting a strictly evaluative approach. This section briefly describes some of these research methods and assesses their utility for this purpose. The use of case studies and other qualitative methods have been selected as being particularly suited to the work of NGOs, and to illustrate applying some practical examples.

3.3.1 Case studies

Case studies may be used in a number of different ways. Across a sufficient number of projects, a random sample of cases can be selected to search for particular characteristics (effectiveness for example) and the findings can then be generalised for the entire population. Similarly, findings for a stratified sample can be generalised for the selected sub-groups within the population. Non-representative samples involving small numbers of case studies or even single cases can be selected to research particular aspects of their content. Some types of selection include extreme or deviant cases, maximum variation cases, critical cases (to disprove a theory for example) or paradigmatic cases (that clearly illustrate or establish a particular school of thought) (Flyvbjerg, 2006).

The INTRAC paper discusses the importance of being methodical and clear in the selection and analysis of case studies, rather than just using stories to illustrate the work of an NGO. One example in the paper is the use of Most Significant Change stories:

Most Significant Change methodology seeks to produce information-rich stories representing the most significant changes brought about by an organisation within pre-specified domains. Within an INGO, the domains can be defined to represent broad areas of change such as strategic objectives or framing indicators. The method of story generation and selection is transparent and replicable, and, if the methodology is followed correctly, stories are verified by a range of different stakeholders before being used. Stories are deliberately acknowledged to be generated through purposeful sampling. A set of stories produced at global level can help summarise the work of an INGO because the rationale for the stories is clear, and the methodology transparent (Simister, 2016)

In practice however, we should take note of critiques of the actual process of gathering and recording stories, including from whom they are gathered and by whom and how representative they actually are (Willets and Crawford, 2007)

ACFID's Education Working Group used a case study approach with pre-specified domains in 2013 to gather evidence of the particular role and value of NGOs in education (Case Study 2), considered in detail below. Members of the ACFID Education Sector Working Group compiled 19 case studies from 11 Australian NGOs and undertook a textual narrative synthesis across the case studies to demonstrate that NGOs bring qualities and capacities to their work in education and development that are unique, efficient, strategic and often underestimated. Both large and small ANGOs participated in providing case studies for the synthesis.

3.3.2 Qualitative methods

Use of other qualitative research methods and tools, such as focus group discussion, semi-structured interviews, participant observation, ethnography and others may be used to research NGO contributions at different scales. These methods undertake primary research. Methods synthesising existing research have already been discussed (under Evaluation Synthesis).

An example of a research approach that used a number of qualitative methods is found in Case Study 3: *Exploring the link between child and youth participation and development effectiveness*. This research project was a partnership between three ANGOs (Transform Aid International, ChildFund Australia, and Live & Learn), their in-country partner organisations, and a university partner, the Institute for Sustainable Futures (ISF) at the University of Technology, Sydney. It ran from 2013 to 2016, and focussed on projects that involved children and youth in a participatory way. These included: youth-led environmental management and livelihood initiatives in Fiji; sport and child clubs in Laos, and child clubs in Nepal.

3.3.3 Strengths and limitations of case study and other qualitative methods

This section focuses on the strengths and limitations of Case Studies 2 and 3.

Being clear on the purpose and questions and making use of the information

Case Study 2 was clearly designed to highlight the particular contribution of NGOs in education and had a very clear purpose. The primary audience was DFAT (then AusAID), targeted by this piece of work to inform ongoing policy dialogue between the Education Sector Working Group and AusAID education personnel. The intent of the synthesis was to demonstrate the unique and necessary contribution of the NGOs to education.

The process was useful for the smaller agencies to see their case studies in print and 'endorsed' by ACFID and AusAID. It benefitted all the participating NGOs in the eyes of AusAID, who now had evidence that the NGOs had something serious to offer in the education sector. The product appears to have had some influence on the 2014 design of the Basic Education Quality and Access in Lao PDR (which includes an NGO consortium component) and, as part of a broader advocacy and influence agenda, it likely influenced the incorporation of Early Childhood Care and Development as the first pillar in DFAT's most recent education strategy, developed in 2014-15. Nonetheless, the project might have benefited from more planning for the strategic use of the document from the beginning of the process.

Case Study 3 was designed to investigate the benefits of a particular NGO development approach, child and youth participation in development initiatives. This was a clear purpose and the team are clear on how they will use the information. Using a research partner brought the particular research skills needed for the project to the partnership.

Applying fit-for-purpose methods in fit-for-purpose ways

Case Study 2 is a very useful example of working across agencies in a particular sector to promote the particular role of NGOs in education. It used a narrative synthesis of case studies with pre-specified domains for the analysis. The co-convenors of the Education Sector Working Group distilled common experience from the case studies as relevant into the framework that had been worked out. The method was straightforward, but fit-for-purpose in that it provided sufficient evidence without using an overly complicated method and was able to be completed in house.

In Case Study 3, different research methods were used for three Learning Areas (the contribution of child and youth participation to development effectiveness, in-country partner organisation outcomes and ANGO program outcomes and child and youth participation in research practice). These included a case study approach where children and youth described 'change stories' that happened because of their own participation; a self-assessment tool; reflection questions (using researcher daily debrief sheets, a post-research survey conducted using an online survey tool, and teleconferences with in-country partners).

Research data was collected for the first phase in 2014 and analysed and synthesised in a Mid Term Learning Paper (2015). The second phase built on the learnings from the first phase and was undertaken in 2016, leading to a final Learning Paper in 2017 that was being finalised at the time of preparing this Case Study. Shorter products will also be developed for advocacy purposes.

Adopting methods and approaches that address complexity

Both Case Studies 2 and 3 used methods that sufficiently addressed complexity. In Case Study 2, some of the case studies drew on material from evaluations, others did not. The quality of the case studies was generally good in terms of evidence, although some persisted in a level of agency branding in spite of the guidelines and one was less substantial in terms of content. It was deemed politically sensible to include all the case

studies however, with the co-convenors deciding which material to use in the synthesis itself, providing a level of quality assurance.

Considering the needs of all partners, including locally

Case Study 2 addressed the needs of the Australian NGOs and DFAT. Local partners were not considered, but it was drawing on secondary data.

In Case Study 3, the involvement of in-country partners and development of their capacity for research was a positive aspect. The Australian partners reflected however that the approach was a bit 'top down' and they should have involved in-country staff earlier and more substantially in the process as well as involving women and youth. They would also seek to use a local in-country University in future.

Capturing the distinctive contribution of the NGOs

Both case studies were specifically designed to capture the distinctive contribution of NGOs. For Case Study 2, this was the purpose of the exercise and the synthesis highlighted six key themes across the sector of education:

- NGOs can reach the unreached.
- NGOs work across the whole education sector but target the gaps.
- NGOs build long-term relationships at community level and strengthen civil society.
- NGOs work locally but have a range of relationships with government up to national level.
- NGOs are an effective, and often efficient, way of bringing about change at system level.
- NGOs' work is technically sound, evidence-based and able to document long-term impact.

Case Study 3 investigated a particular development approach being used by the NGOs, youth participation, and found evidence of its links to development effectiveness.

Not being overly complicated, technocratic or exacerbating fragmentation

Case Study 2 was based on secondary research - the case studies provided by the participating NGOs - and did not exacerbate fragmentation. It was not overly complicated or technocratic.

Case Study 3 involved primary research. The methods took some time to refine to make fit-for-purpose. This Case Study demonstrates both the benefits and the complexities of working in a partnership across three medium and small NGOs and a University partner to understand more about a particular development approach. The model could be streamlined based on the reflections of the partners.

Having adequate resourcing for the purpose

The process for Case Study 2, from conception to publication, took about a year in 2013, led by the two co-convenors of the Education Sector Working Group, who, with two others, formed a core group of four representatives from four NGOs. The core group managed and implemented the whole process, with encouragement and assistance with final formatting and printing from ACFID and funds for publication (170 hard copies) provided by seven of the participating NGOs as well as a small ACFID contribution.

Coordination was a time-consuming process, firstly getting the case studies from the participating NGOs and then among the core team working out how to present the synthesis, which is in two parts – the synthesis itself followed by each of the case studies. The main

work of the narrative synthesis was undertaken by the two co-convenors. The amount of additional work for the core group, especially the co-convenors, was considerable. Resourcing such projects is a big consideration, where they need to be considered part of the 'day job' and not an added extra.

In Case Study 3, funding was insufficient for a more collaborative process with local partners. In fact, the activity over-reached on the resources available, which included the University partner investing significantly. Both a budget and a commitment to contribute not only to each organisation's own data collection activities, but to the wider budget requirements was needed from the ANGO partners.

Research approaches and methods, working with both primary and secondary data, are appropriate if the purpose of capturing, integrating and reporting across scales is to gain further knowledge, understanding and insight as opposed to assessment or judgment purposes (when evaluative approaches are used). The methods however are common to both approaches. The two research case studies had clear purposes and both were targeted at understanding or demonstrating distinctive contributions of NGOs. Both highlighted the need to have sufficient resources for the purpose, and both underestimated the time it takes to work collaboratively. Smaller NGOs were involved in both these case studies and valued the opportunity to demonstrate their contributions.

3.4 Combining the methods

Applying a common performance framework at an agency level or at a program level is a way to organise monitoring and evaluation processes and products. There are numerous ways of doing this, and many of the different methods and tools discussed above can be applied in combination to generate information across different scales. This section looks at two types of models: donor-led program approaches and agency-centred approaches

3.4.1 Donor-led program approaches

In this model, the donor designs a program in accordance with certain principles, which extend to its monitoring and evaluation system.

Case Study 4 looks in detail at the monitoring, evaluation and learning framework for the Australia Africa Community Engagement Scheme (AACES), a partnership between DFAT, ten Australian NGOs and their in-country partners across 11 Sub-Saharan African countries. The ten Australian NGOs (ActionAid, The Australian Foundation for the Peoples of Asia and Pacific Ltd, Anglican Overseas Aid, CARE Australia, Caritas Australia, Marie Stopes International Australia, OXFAM Australia, Plan International Australia, WaterAid Australia and World Vision Australia) delivered community-based interventions over a five year period (2011 – 2016), across the three sectors of water and sanitation, food security and maternal and child health (see Annex 3).

3.4.2 Agency-led approaches

Australian NGOs might design an overarching performance framework to be able to measure agency performance, beyond the level of individual projects and initiatives. These can be used for high level accountability (for example to the Board or to donors), for fundraising and public relations or for learning and performance improvement purposes. They might apply different lenses (sector or country for example) in order to consider the performance information at those levels, and/or apply a framework that applies to the whole agency.

Case Study 5 looks in some detail at *OXFAM Australia's approach to reporting against its Strategic Plan 2014-19* (Annex 3). A common analysis framework which guides a suite of products includes the Strategic Plans goals and results and a set of strategic evaluative

questions. These questions are addressed through the suite of products and culminate in a Strategic Plan Report.

3.4.3 Considerations in combining methods

This section draws on the insights from Case Studies 4 and 5 to make some observations about the benefits and challenges of applying the methods in combination.

Being clear on the purpose and questions and making use of the information

In Case Study 4, the purpose of the program level monitoring and evaluation was to meet DFAT's accountability requirements but also to strengthen the program to target and serve the needs of marginalised people through learning, collaboration and exchange among AACES NGOs and between AACES NGOs and DFAT. This was a clear purpose.

The program configuration (10 Australian NGOs and their in-country partners across 11 Sub-Saharan African countries and 3 sectors) was, however, an 'administrative' construct created by the DFAT funding window; there was no developmental rationale that bound together these countries, sectors or agencies. A focus on monitoring and evaluation at the level of a single country, development approach or sector would be more developmentally coherent.

The purpose of OXFAM Australia's approach in Case Study 5 is also clear - to monitor the overall performance of OXFAM Australia programs, contribute to thematic strategy and program learning and development and support engagement with institutional donors, supporters and the public. Individual products within the suite are structured so that they address strategic learning questions:

Outcomes

- What positive and negative changes have OXFAM Australia supported programs contributed to against prioritised Strategic Plan goals and results?
- Who has benefited and how have they benefited? What benefits have marginalised groups experienced?
- Are the changes at local, regional, national or global levels?

Effectiveness

- What strategies are effective/not as effective in enabling change at different levels in programs and projects?
- What strategies and tactics have worked in supporting people, communities and civil society to influence decision makers? How effectively have marginalised people been supported to take leadership roles and influence decision makers?
- How effectively has OXFAM worked with implementing and strategic partners?
- What has been OXFAM's contribution to program/project outcomes and effectiveness?

The individual products are to be provided at specific points in the OXFAM business development cycle to contribute to program development and decision making about the portfolio. The OXFAM Australia approach will continue to emerge and develop. It takes a learning approach, with continuous learning loops, but there is still work to be done to align the individual pieces of analysis to their use. It is a big organisation with a dispersed focus, making it difficult to find the right moment when particular reporting will help inform particular programming decisions in the business cycle, such as budgets, country engagement strategies, project designs and monitoring, learning and evaluation approaches or the development of thematic strategies. Timing is one challenge but also the specific messaging of the products is another.

Applying fit-for-purpose methods in fit-for purpose ways

In Case Study 4, the key features of the program-level monitoring included:

1. Annual Review Workshops.
2. Annual Program Report
3. Peer Review
4. Mid-term and Final Reviews
5. Value for Money Assessment

In practice, the key features of annual reporting across the whole program were aggregating common indicators and textual narrative synthesis across individual case studies. Common cross-cutting themes were extracted from the case studies provided by the NGO partners to broadly illustrate some of the commonalities, for example in mobilisation and organisation of men and women in communities, facilitating service delivery to poor and marginalised people, enhancing social inclusion and gender equality, capacity development and social accountability. To some degree this demonstrated the unique contribution of the community-based NGO approach, but coherently communicating the breadth and diversity of this program remained a challenge.

In Case Study 5, the aim is to report every two years against the Strategic Plan through a Strategic Plan Reporting Summary Assessment, drawing on a suite of products:

1. OXFAM Australia's Portfolio Analysis
2. OXFAM Australia Evaluation Summary
3. OXFAM Australia Evaluation Reviews
4. OXFAM Australia Program Performance Reviews
5. OXFAM Thematic Frameworks
6. OXFAM Australia Strategic Outcomes Analysis

The products use many of the methods already described. For example OXFAM Australia's Portfolio Analysis draws heavily on aggregated indicators. OXFAM Australia Evaluation Reviews are desk-based narrative syntheses of program evaluations. An identified risk relates to the quality of program evaluation documentation.

Each of the products needs to be fit for purpose and available at the right moment to the right audience and this requires identifying the right connections within OXFAM Australia's business processes.

Adopting methods and approaches that address complexity

The approach to monitoring and evaluation in Case Study 4 did not attempt to oversimplify the complexity of the development challenges being faced in the eleven countries nor the complexity of the program itself. The non-prescriptive approach to monitoring and evaluation systems was key as well as the development of agreed outcome statements. Trying to find coherence in monitoring and evaluating a disparate program across multiple sectors, multiple countries and multiple agencies was a major challenge, with the added complexity of being driven from the ground up, in other words the design was not prescribed. A focus on cross-cutting issues such as gender and disability was a helpful strategy for managing aspects of the complexity.

OXFAM's performance system is designed overall to embrace complexity and not be reductionist.

Considering the needs of all partners, including locally

In Case Study 4, the monitoring and evaluation was heavily centred close to the action in Africa and included the in-Africa NGO partners. The annual workshops to reflect on the program, allowed in-Africa partners to share in the review, benefit from capacity building and participate in the management of the program. At a country level, the partnership approach helped local NGOs to collectively advocate with the national governments, for example, around water programming in Tanzania and budgeting for water in Malawi.

OXFAM's performance system is aligned directly to OXFAM's Strategic Plan and should assist OXFAM in performing its mission and vision and directly help with advocacy efforts for example. In accountability terms the system will meet DFAT's requirements. Accountability to partners and communities is variable and more complex, because OXFAM International manages programs in-country (although OXFAM Australia is involved in governance of OXFAM International and its monitoring, evaluation and learning processes and in supporting country operations where it holds the contract).

Capturing the distinctive contribution of the NGOs

Particular elements of the community-based NGO approach in Case Study 4 were captured through the mix of M&E methods, at project and program level, with common cross-cutting themes around community mobilisation, reaching poor and marginalised people, enhancing social inclusion and gender equality, capacity development and social accountability. Case Study 5 is designed around OXFAM's Strategic Plan, directed at its distinctive contribution.

Not being overly complicated, technocratic or exacerbating fragmentation

For Case Study 4, the individual NGO projects and their M&E systems as well as the overarching program M&E system were all designed as part of the program design. It built from existing NGO M&E systems so was careful not to impose additional requirements. The approach was necessarily intense as it was dealing with such a diversity of countries, NGOs, in-country partners and sectors as well as DFAT as the funding partner. It was designed to be as strategic as possible and actively avoided a technocratic approach.

OXFAM's reporting system appears quite complex, but as yet untested in practice.

Having adequate resourcing for the purpose

In Case Study 4, Monitoring and evaluation was identified in the initial design as an important aspect of this program. NGOs were encouraged to use up to 10% of their budget for monitoring, evaluation and research activities, and up to 10% of the overall AACES budget had similarly been allocated towards management, monitoring, evaluation and research. The development of monitoring and evaluation approaches in the NGOs and with partners, the annual reviews and meetings, engagement in the PSC, reporting on projects and across the program required a considerable investment of NGO and DFAT staff time, including Canberra-based staff and a full time position based in Nairobi as well as consulting costs. Allowances for these activities were provided in the budget.

Two OXFAM Australia staff are assigned to the implementation of Case Study 5, one for the portfolio analysis and the other (who has other responsibilities) working closely with internal and external experts to implement the remainder. Maintaining this level of staffing and sufficient budget to implement will be important.

The monitoring, evaluation and learning approach adopted by Case Study 4 (AACES) was highly valued by all the partners. It was distinctive in its localised approach, the dedicated and adequate resourcing of monitoring and evaluation and its embracing of complexity. Case Study 5 (OXFAM Australia's reporting system) is designed around answering a set of strategic learning questions and also embraces complexity. These two examples illustrate how the methods described in this report can be combined and applied in practice across the 'administrative' scales of a donor funding window and across an agency.

4 Conclusions and Recommendations

The purpose of this research is to show how evidence of outcomes and impact can be better captured, integrated and reported on across different scales for Australian NGOs working in international development. The research has drawn extensively on the literature (both

academic and 'grey'), studied five cases in detail, considered numerous other examples in less detail, and consulted with Australian NGOs, academics and monitoring and evaluation consultants. The research is not a comprehensive treatise on the subject, rather it focuses on describing and analysing some of the major approaches and methods that are likely to be relevant for Australian NGOs and draws on numerous examples, including the five case studies, to illustrate.

This iterative and participatory process developed, tested and refined seven factors that should be considered in any exercise involving Australian NGOs to capture, integrate and report on outcomes and impact across different scales.

Factors to consider when capturing, integrating and reporting on outcomes and impact across different scales

- Being clear on the purpose and questions and making use of the information
- Applying fit-for-purpose methods in fit-for purpose ways
- Adopting methods and approaches that address complexity
- Considering the needs of all partners, including locally
- Capturing the distinctive contribution of the NGOs
- Not being overly complicated, technocratic or exacerbating fragmentation
- Having adequate resourcing to meet the purpose

In applying these factors across the three main methodological approaches – using indicators, evaluative approaches and research approaches, some of the strengths and limitations in the suitability and application of these approaches were highlighted, discussed below. This chapter concludes on the whole of the research under those headings.

4.1.1 Being clear on the purpose and questions and making use of the information

Some NGOs are satisfied that making their contribution one project at a time is sufficient and there is nothing to be gained from investigating what it all adds up to. Roger Riddell however sees the 'absolute and comparative advantage' of the NGOs in their wider and long term impact (Riddell, 2013: 385) which requires thinking beyond single projects.

Selecting and implementing the most appropriate methods or combinations of methods is both a technical and political issue. NGOs have a tendency is to think of their own agencies, reputations and public relations above the need for local engagement, collaboration and collective thinking about the contributions of NGOs and partners to achieving development impact and the SDGs. In part, the NGOs are operating in a competitive funding environment which is not conducive to collaboration. In part many operate from a project mindset that detracts from thinking at more strategic levels. If the NGOs are interested in genuinely deepening their understanding of their contribution to development and achieving the SDGs and are willing to work collaboratively with others, then a more strategic approach to monitoring, evaluation and learning at and across different scales is appropriate.

Any exercise in capturing, integrating and reporting across different scales of work must begin with a very clear purpose and questions. Part of thinking through the purpose includes defining the scales. Such scales might usefully be defined by what makes sense developmentally, for example across sector, country or global levels, as opposed to more 'administrative' scales of single donors or single NGOs. The information generated must be useful and used – for example in demonstrating impact, advocacy, learning, decision-making or accountability.

All five case studies had a clear purpose which covered a variety of scales. One of the learnings from those case studies that operated across different countries was the absolute need to have a deep understanding of the context and culture of the individual countries involved.

4.1.2 Applying fit-for-purpose methods in fit-for-purpose ways

Form follows function and it is the purpose of the exercise in capturing, integrating and reporting across different scales of work that determines the choice of method or methods. At a broad level, if the purpose involves judgement then evaluative approaches and methods should be used, if the purpose is primarily about deepening understanding then research approaches are suitable and if the purpose is about managing a portfolio then using indicators might be applied (Table 3). Noting however that evaluation is a type of research and the methods used are largely common. Table 3 is an oversimplified representation as there is overlap and blurring of lines between the approaches. Using indicators is best done as part of a performance story or in combination with evaluative or research approaches. All three approaches can be used to demonstrate results.

TABLE 3: SELECTING APPROACHES ACCORDING TO PURPOSE

Using indicators	Evaluative approaches	Research approaches
When the purpose is to manage the basics of a portfolio of projects or to demonstrate results. Should avoid using in isolation.	When the purpose involves judgements, usually about effectiveness, efficiency and/or appropriateness	When the purpose involves gaining knowledge, understanding and insight

Quantifying outcomes and impact in both quantitative and qualitative terms is very important. But the popularity of an agency-centric use of aggregating indicators among international NGOs is a cause for some unease, because on their own they do not address complexity, are often not considering the needs of local partners and are not helping to further a global understanding of the distinctive contribution of NGOs. One of the issues is the isolated use of indicators in reporting. They are better used as part of telling a ‘performance story’. The main story line of a performance story is how well a project or program has performed in relation to what was expected and what will now be done differently to better ensure future performance. This idea of performance stories was designed with single interventions in mind, but it contains an important principle of not only reporting the outcomes but also discussing the evidence available to demonstrate the contribution made by the intervention to those outcomes. This principle can also be applied at an aggregate level if the use of indicators is integrated into a narrative, or into evaluative or research approaches.

There are basically two ways that evaluative methods can be used to help capture, integrate and report on outcomes and impacts across different scales of work. The first is to undertake a strategic evaluation that targets the particular scale or scales rather than a single project. The second is to synthesise information available in existing project-level evaluations.

Using a strategic evaluations to evaluate across different scales is a complex exercise that can be methodologically challenging and expensive. However, if the purpose and evaluative questions warrant such an exercise and align with methods that address complexity (see below) it can provide explanatory insight into outcomes and impact at higher orders of scale than individual projects. Guides are available (such as the BOND ‘Choosing Appropriate Evaluation Methods Tool’) for selecting evaluative methods according to the types of questions being asked.

A strategic evaluation, as in Case Study 1, will often draw on existing documentation and evaluations as well as complement this with primary research. In this sense, evaluation syntheses are not always applied as standalone methods but may be integrated into ‘strategic’ evaluation designs.

Realist synthesis and other textual narrative syntheses potentially address complexity and can make good use of existing evaluative information. The exercises require careful and

transparent methods. Because they draw on existing evaluations there is likely a time and cost saving compared to a 'strategic' evaluation. However they are also dependent on the quality of the underlying information.

Research approaches and methods, working with both primary and secondary data, are appropriate if the purpose of capturing, integrating and reporting across scales is to gain further knowledge, understanding and insight as opposed to assessment or judgment purposes (when evaluative approaches are used). The methods however are common to both approaches. Case Studies 2 and 3 had clear purposes and both were targeted at understanding or demonstrating distinctive contributions of NGOs. Both highlighted the need to have sufficient resources for the purpose, and both underestimated the time it takes to work collaboratively. Smaller NGOs were involved in both these case studies and valued the opportunity to demonstrate their contributions.

Within these broad approaches, the most fit-for-purpose method or combinations of methods need to be selected in accordance with the particular questions being asked. These should be the minimum required for the purpose, demonstrated through Case Study 2, which used a straightforward narrative analysis of case studies to demonstrate the particular contribution of NGOs to education. Other case studies, such as Case Study 1, reflected that they had underdone the methodology and proposed amendments for future studies. Case studies 2 and 3 are useful examples of how the work of smaller NGOs can be captured in a meaningful way. Case studies 4 and 5 illustrate how methods can be applied in combination.

4.1.3 Adopting methods and approaches that address complexity

Approaches used for integrating and reporting across different scales of work also need to address complexity or they risk perpetuating over-simplified understandings of program logics and development and will not be useful. This is made more difficult by the fact that existing M&E systems (usually at a project level) have not yet responded to more recent understandings of complexity, which require quite different approaches to those commonly used. For example, 'thinking and working politically' suggests fewer pre-defined results, mixed approaches to monitoring and evaluation and the need for enhanced collaboration.

Some evaluative and research methods can address complexity while aggregating indicators generally does not if applied on its own (Table 4). Mixed methods, combining quantitative and qualitative approaches are preferred. Some qualitative methods that address complexity include ethnography, contribution analysis, realist evaluation, an adapted panel approach (adapted from Case Study 1), realist synthesis and other qualitative methods.

TABLE 4: ADOPTING METHODS AND APPROACHES THAT ADDRESS COMPLEXITY

Using indicators	Evaluative approaches	Research approaches
Generally does not address complexity but can be combined with other methods	Some methods – e.g. ethnography, contribution analysis, realist evaluation, an adapted panel approach, realist synthesis and other qualitative methods – address complexity	

4.1.4 Considering the needs of all partners, including locally

NGOs seem to have not concentrated heavily on strengthening local monitoring and evaluation systems, for example to strengthen partner civil society systems to operate within their own regulatory environments. While international development NGOs have been involved in helping both governments and NGOs develop their own civil society regulatory systems, they have been more focused on strengthening their own monitoring and evaluation systems than those of their in-country partners. Similarly international NGOs have been less concerned with strengthening the capacity of civil society to monitor government performance and achievement of the SDGs. This seems inconsistent with the commitment of

ANGOs to build capacity within communities, civil society and government to address their own development priorities.

The needs of all partners need to be balanced. At the moment, the tendency is for the monitoring, evaluation and learning needs of donors and implementing agencies including NGOs to take priority over those of local partners and governments.

None of the case studies or other examples in this report began from the perspective of local ownership or alignment with local systems. They were all 'owned' by donors or ANGOs. Two of the case studies made serious efforts to work with local partners – Case Study 3 and Case Study 4 – and this may have resulted in some sustainable benefits among the partners involved. However, the absence of consideration about what the needs of the partners might be in terms of monitoring, evaluating and learning across scales is perhaps a consequence of the competitive funding environment, a project mindset or that of agency self-preservation discussed above. The reporting of the SDGs at a country level may provide more impetus for a shift to local perspectives and equitable partnerships.

4.1.5 Capturing the distinctive contribution of the NGOs

The significance of the NGOs is in their distinctive contribution and the impact of their collective or collaborative effects, but their monitoring and evaluation systems are often narrowly based. Some NGOs have been historically progressive in promoting alternative development paradigms over the mainstream linear theories of economic development. The diverse roles of NGOs and CSOs now numbering in their tens of thousands should be able to help inform the broader aid community on the effectiveness of complex aid mechanisms, particularly in addressing human rights, inequality and power relationships. The degree to which the distinctive contribution of the NGOs is captured largely depends on the purpose and types of questions the exercise is seeking to address.

Case Study 1 was specifically looking at the broader impact of NGOs and CSOs. Case Study 3 was investigating the relationship between child and youth participation and development effectiveness. Seeking more information about how NGO development approaches promote gender equality, human rights, empowerment and participation can be a useful contribution of NGOs to the broader development agenda and help recognise their particular contribution.

An important consideration in capturing the distinctive contribution of NGOs is a focus on the nature of the intervention and how this has contributed to outcomes and impact. This requires a mix of methods and approaches that not only focus on the outcomes and impact but also include some explanatory power. Many of the research and evaluative approaches described, such as realist evaluation and realist synthesis, seek to understand the mechanisms of how the particular change has come about.

4.1.6 Not being overly complicated, technocratic or exacerbating fragmentation

The mechanical use of performance information for reasons of accountability and control (Vähämäki et al., 2011, Eyben, 2015) and the 'rising tide of technocracy' (Banks et al., 2015: 708) are plaguing monitoring and evaluation systems. When capturing, integrating and reporting across different scales there is a risk that these characteristics will be reinforced, so a conscious effort to identify useful measures is needed. Aggregating indicators carries a particular risk, identified in a UK study of monitoring, evaluation and learning systems, which is that quantitative data is easier to store and analyse than qualitative data (ITAD, 2014). It must be the purpose that drives the choice of method and not the means of collecting the data.

Case studies 2 and 5 made use of existing performance information and did not exacerbate fragmentation. Case study 1 used a mixture of primary and secondary research. It cut through technocracy by focusing on higher order impacts. Case Study 3 introduced new

research tools to gather primary data on the relationship between child and youth participation and development effectiveness. The process could be more streamlined in future iterations. The program level M&E system for Case Study 4 (AACES) was part of the program design itself. The involvement of local partners in developing meaningful measures and the reliance on existing NGO systems mitigated against technocracy and fragmentation.

4.1.7 Having adequate resourcing to meet the purpose

Most of the case studies discuss the adequacy of resources as a major practical factor in their implementation, in terms of staff time allocated as well as funds. In Case Study 4 (AACES), NGOs were encouraged to use up to 10% of their budget for monitoring, evaluation and research activities, and up to 10% of the overall AACES budget had similarly been allocated towards management, monitoring, evaluation and research. This meant that the NGOs and DFAT were able to allocate the necessary time to generate and analyse information for program level monitoring. Other case studies were under-resourced – Case Study 3 (Child and Youth Participation) over reached on their budget and Case Study 2 (Reach, Relationships and Results) reflected on the amount of work involved in the partnership and how this was ‘additional’ to regular jobs. The staff time needed for collaborating and coordinating within agencies and across agencies should not be underestimated. Resources may also be needed for buying in particular methodological expertise or for building capacity of staff or partners to undertake the synthesis. Finally, the time frame for exercises in integrating and reporting on outcomes and impact across different scales of work are likely to stretch over a year or more.

It takes time and money to build capacity across partners and there are political challenges in obtaining such investment especially as it sits outside of project funding. The competency and availability of ANGO staff or in-country partner staff is an issue. The capacity for this type of work is less likely to exist at the local level where the focus is on implementation. The implications for workload and allocating staff for the task also need to be prioritised, which is difficult to do over project implementation.

Pooling of resources between agencies was involved in Case studies 2 and 3. Although this brings its own administrative and other issues, it is a useful way to increase the resources available for a collaborative exercise and to enable smaller NGOs to participate. Smaller NGOs might see their advantage in their ability to leverage and to be nimble and flexible compared to the larger agencies. In this context there may be possibilities for the smaller NGOs to identify opportunities for adding strategic value to collaborations with larger NGOs.

There needs to be a very clear purpose to investing sufficient resources in such an undertaking. All of the case studies demonstrated that this is a resource intensive exercise, with only Case Study 4 demonstrating an adequate budget for the task. Working collaboratively consumes considerable resources. In addition, the methodological challenges in better capturing, integrating and reporting on evidence of outcomes and impact across different scales should not to be underestimated. The technical complexities mean that involvement of specialist evaluators and researchers is likely to hold some benefits, in addition to building those capacities among ANGOs and their local partners.

4.2 Recommendations

Capturing, integrating and reporting on outcomes and impact across different scales in this research is about a more strategic approach to measuring development effectiveness by ANGOs and requires high level strategic intent by the organisation, with associated commitment, processes and capabilities. It involves both political and technical challenges. In some ways, starting with monitoring, evaluation and learning across different scales is premature because it makes most sense to do this when agencies are *programming* across different scales. However, there is a chicken-and-egg element to this – if the benefits of

capturing, integrating and reporting across different scale are visible and obvious, it can help drive programming behaviours.

This research is the first piece of a larger body of work over the next few years, when ACFID and the RDI Network will – resources permitting – use the research findings to inform the development of other pieces of work. The first step after finalising this report is to hold co-designing workshops with monitoring, evaluation and learning practitioners in 2017 to unpack the findings of the research. Three recommendations are designed to further progress the use of this report and facilitate future action around how evidence of outcomes and impact can be better captured, integrated and reported on across different scales of work:

- 1. Agencies should consider adopting a more strategic approach to monitoring, evaluation and learning and consider the seven factors in any exercise involving Australian NGOs to capture, integrate and report on outcomes and impact across different scales.**
- 2. ACFID or RDI Network should establish a resourced community of practice for monitoring, evaluation and learning practitioners. This group should take forward this report by:**
 - a. Sharing experience and expertise about monitoring, evaluation and learning, particularly around methods that address complexity**
 - b. Capturing lessons, sharing insights and supporting existing or proposed member initiatives to learn from and demonstrate the value of capturing, integrating and reporting on outcomes and impact across different scales**
- 3. DFAT should consider ways to support good practice in capturing, integrating and reporting on outcomes and impact across different scales by ANGOs, including in partnerships and collaborations. These ways might include designing its support to ANGOs in ways that support collaboration rather than competition, and recognising the full costs of monitoring, evaluation and learning.**

Annex 1: Research Methodology Summary

The purpose of this research is to show how evidence of outcomes and impact can be better captured, integrated and reported on across different scales of work.

Key Questions

The research addresses the following questions (questions 6 and 7 to a lesser extent in accordance with the TOR):

1. What are the different scales of work that are relevant to this study from the perspective of ACFID member agencies, ACFID and DFAT?
2. Which factors should inform the selection, application and critical assessment of practices for capturing, integrating and reporting on evidence of outcomes and impact across these scales?
3. What approaches, methods and tools are available to capture, integrate and report on evidence of outcomes and impact across these different scales?
4. Taking account of the principles referred to above, what are the strengths and weaknesses of each of these approaches, methods and tools when applied at these different scales?
5. What issues emerge – both political and technical - when applying these approaches, methods and tools at different scales and how can they be addressed?
6. What factors are helping/hindering the take up and principled use of these approaches, methods and tools at different scales by individuals, by agencies, by groups of agencies and by donors?
7. What can agencies, ACFID, RDI Network and DFAT do to improve performance in this area?

A pragmatic paradigm to match the purpose and the short time frame guided this research, which relied on qualitative methods designed to answer each of the research questions in ways that will be useful for the ACFID membership. Five case studies were selected to illustrate the various methods and approaches in practice. The approach was participatory, with the views of experts and practitioners forming part of the data, and iterative, with each step of case study production and report development built on the last, on the basis of review by experts and practitioners.

Research Question	Approach	Product
1. What are the different scales of work that are relevant to this study from the perspective of ACFID member agencies, ACFID and DFAT?	Use an iterative process with the steering committee to establish an initial framework of scales and principles. The initial framework will be iteratively modified as the research progresses.	Implementation plan Initial framework of scales and principles
2. Which principles should inform the selection, application and critical assessment of practices for capturing, integrating and reporting on evidence of outcomes and impact across these scales?	Agree implementation plan with steering committee, including clarifying the nature of the outputs.	
3. What approaches, methods and tools are available to	Literature review, to include the academic evaluation literature,	

capture, integrate and report on evidence of outcomes and impact across these different scales?	<p>selected donor agency and NGO publications and websites.</p> <p>Key informant communications with a limited number of M&E specialists/ practitioners</p> <p>Initial sharpening of meaning of 'approach' to discuss with Steering Committee</p> <p>Descriptive study of 5 selected case studies and interviews (where possible) with implementing practitioners</p>	
4. Taking account of the principles referred to above, what are the strengths and weaknesses of each of these approaches, methods and tools when applied at these different scales?	Drawing on the principles, literature, current practice, group the methods and analyse against the agreed principles. Feedback sought from the steering committee on a draft working paper resulting from this exercise.	Working paper
5. What issues emerge – both political and technical - when applying these approaches, methods and tools at different scales and how can they be addressed?	Half-day workshop with project stakeholders to contest the findings of the analysis, identify further issues of application and possible solutions (draft recommendations). Presentation and feedback from DPC	Workshop DPC meeting Draft report
6. What factors are helping/hindering the take up and principled use of these approaches, methods and tools at different scales by individuals, by agencies, by groups of agencies and by donors?	Review of the draft report to the steering committee will specifically seek additional comment on this question.	Final report
7. What can agencies, ACFID, RDI and DFAT do to improve performance in this area?	Response to the final recommendations in the final report by the steering committee	Dissemination at workshop/ masterclass

Limitations

The time constraints limit how comprehensive this research can be. Not every approach and method was identified and, as identified in the TOR, any discussion of tools is limited. However, the intent is to identify the main approaches of relevance to ACFID members. The research was also limited by the availability of the Steering Committee and key stakeholders, as the methodology was heavily reliant on their input.

Annex 2: Istanbul Principles for CSO Development Effectiveness

1. Respect and promote human rights and social justice

CSOs are effective as development actors when they ... develop and implement strategies, activities and practices that promote individual and collective human rights, including the right to development, with dignity, decent work, social justice and equity for all people.

2. Embody gender equality and equity while promoting women and girl's rights

CSOs are effective as development actors when they ... promote and practice development cooperation embodying gender equity, reflecting women's concerns and experience, while supporting women's efforts to realize their individual and collective rights, participating as fully empowered actors in the development process.

3. Focus on people's empowerment, democratic ownership and participation

CSOs are effective as development actors when they ... support the empowerment and inclusive participation of people to expand their democratic ownership over policies and development initiatives that affect their lives, with an emphasis on the poor and marginalized.

4. Promote Environmental Sustainability

CSOs are effective as development actors when they ... develop and implement priorities and approaches that promote environmental sustainability for present and future generations, including urgent responses to climate crises, with specific attention to the socio-economic, cultural and indigenous conditions for ecological integrity and justice.

5. Practice transparency and accountability

CSOs are effective as development actors when they ... demonstrate a sustained organizational commitment to transparency, multiple accountability, and integrity in their internal operations.

6. Pursue equitable partnerships and solidarity

CSOs are effective as development actors when they ... commit to transparent relationships with CSOs and other development actors, freely and as equals, based on shared development goals and values, mutual respect, trust, organizational autonomy, long-term accompaniment, solidarity and global citizenship.

7. Create and share knowledge and commit to mutual learning

CSOs are effective as development actors when they ... enhance the ways they learn from their experience, from other CSOs and development actors, integrating evidence from development practice and results, including the knowledge and wisdom of local and indigenous communities, strengthening innovation and their vision for the future they would like to see.

8. Commit to realizing positive sustainable change

CSOs are effective as development actors when they ... collaborate to realize sustainable outcomes and impacts of their development actions, focusing on results and conditions for lasting change for people, with special emphasis on poor and marginalized populations, ensuring an enduring legacy for present and future generations.

Annex 3: Case Studies

CASE STUDY 1: Tracking Impact: An exploratory study of the wider effects of Norwegian civil society support to countries in the South

Norway channels a significant proportion of its official aid (over 20%) through Norwegian NGOs and wanted to know more about the overall and wider impact of NGOs and CSOs on development and their long-term contribution to poverty reduction. NORAD's Civil Society Department was searching for a new approach because traditional evaluation approaches were unable to answer these questions. In 2011, they established a Civil Society Panel to give 'an informed opinion and judgement about Norwegian CSO performance' and in particular with 'examining and assessing the wider effects of Norwegian civil-society interventions in four countries – Ethiopia, Malawi, Nepal and Vietnam – and drawing overarching conclusions on wider impact'. The study itself is published (Abuom et al., 2012) as well as an analysis of the approach used, by one of the Panel members (Riddell, 2013). This case study draws on both sources.

The Panel performed the following tasks:

- Synthesised what was known about results based on a rapid review of existing studies and evaluations, and identified gaps in data and information.
- Explored and discussed the wider effects of Norwegian civil society interventions in the four case-study countries.
- Reflected on the lessons learnt and made a series of recommendations to NORAD and the CSOs on the wider and longer-term impact of CSO interventions.

The Panel grouped what they called 'development results' in three categories: Improved quality of life for individuals and communities; wider and long term effects; and strengthened organisational capacity. They developed 19 working hypotheses and 56 questions to guide the assessment. Because the first category is often the subject of other work, the focus was more on the second and thirds categories.

Each country visit lasted one week. It involved meetings and interviews with Norwegian CSOs and their partners, with government representatives, independent researchers and journalists, and with other International NGOs and other donors. The country work also involved reviewing documents, reports, studies and evaluations, especially those shedding light on the wider and long-term impact questions. Workshops with key local CSOs also took place, where the relevance and validity of the proposed hypotheses were discussed.

The Panel recognised the difficulties of the task, including the great variety of activities undertaken by the CSOs; the difficulties of assessing outcomes like 'helping give a voice to and empower poor people'; and problems of attribution and the absence of the counterfactual. They therefore referred to the exercise as an 'exploratory study' rather than an evaluation or an impact assessment.

Summary of the main findings

This section does not intend to duplicate all the findings documented in the report, but to focus on the wider impact uncovered by the assessment, and the methodological issues in uncovering it. As anticipated, significant and tangible results of projects funded by Norwegian CSOs were documented at the individual and community level, and in general individuals and communities benefited from direct and indirect support in areas such as

health, education, micro-credit and agriculture, though mostly not in very large numbers. More surprisingly, the data and evidence at project level was found to be quite weak.

The assessment found ample evidence of wider effects of Norwegian-funded CSOs, but a serious lack of information about them.

'In spite of the lack of robust and easily accessible documentation on wider effects, the Panel was able to find plenty of examples of the wider effects that some projects were having, and this evidence was found in every one of the four case-study countries. Specific examples of wider effects encompassed each of the following areas: (a) replicated, scaled-up and innovative initiatives; (b) the monitoring of government programmes in order to hold the government accountable; and (c) influencing legislation and changing policy processes. Importantly, too, as noted above, CSOs continue to have a wider impact through the significant contribution they make to the overall provision of especially health and educational services in many countries.

However, in part because CSOs were not often asked or challenged to analyse the wider impact of their work, robust and systematic evidence was hard to find. What is more, we found numerous instances where CSOs gave examples of changes in laws or policies in areas where they had done advocacy, but they were unable to explain precisely what effect their activities and actions had had on the changes that had occurred. Most examples found are in the area of replication and issue-based policy advocacy; CSOs involved in preparing a new national policy on disability, child rights and HIV/AIDS either by being members of working groups or advocating for change from the outside.

Though we found examples of innovation and replication, the innovative profile was relatively weak: most projects used well-known approaches and technologies and what was perceived as innovative had often been tried out elsewhere.

Our assessment of the nature of impact is broadly consistent with the recent CIVICUS survey, namely that civil society achieves the highest level of impact in the social sector, but less political impact, including influence on policy making (see CIVICUS 2011:36).' (Abuom et al., 2012: 8).

In Riddell's analysis he highlights the issues that were faced in assessing the wider impact of CSO interventions. One major observation he makes of his assessment that wider impact was achieved more by chance than design:

'This is because the overwhelming majority of CSOs predominantly view their work and make decisions on how to engage in the development process through the prism of individual projects: they neither assess their interventions nor decide what they should or might do from a wider perspective, working out how they could or might make a bigger difference.' (Riddell, 2013: 279).

On the third category, strengthening civil society, the Panel struggled with the assessment because it required a common understanding of the core role of civil society, with three competing perspectives: the first to do with advancing development and reducing poverty, the second to do with strengthening democracy and governance and the third that strengthening civil society is an end in itself. Even while maintaining these competing views, the Panel found evidence of building the capacity of individual organisations, usually in areas such as monitoring, that would help them to comply with the Norwegian NGO and NORAD's requirements. There were some examples of a less narrow focus and successes in building sustainable and viable organisations.

Reflections

A major point of reflection by the Panel was the importance of local context, including culture and history, in making assessments of individual projects but even more so in assessing wider NGO impact.

This underlines the need to have an in-depth understanding of each country, the nature of civil society, how it has evolved and how it is perceived by different stakeholders. This provides the backdrop for understanding where civil society might best contribute to development, poverty alleviation and peace building, how best it might help to monitor the actions of a government and its agencies, how best it might help to monitor human rights abuses and work for the fulfilment of basic rights or how civil society might be strengthened and what the implications might be for the wider civil society when assistance is provided to particular CSOs or groups of CSOs. (Abuom et al., 2012: 16)

Riddell further documents the Panel's reflections around the findings and the process of the review. On methods, Riddell stresses that this 'was termed an 'exploratory study' in order to make clear that this was not an evaluation as traditionally understood, nor was it conceived of as an alternative to more rigorous assessments and evaluations. It aimed to do something different, building on the insights and approaches of both more orthodox assessments and research.' (Riddell, 2013: 386).

The main reasons for adopting this approach was that no other approach was sufficiently relevant for studying the wider effects of NGO impact across a whole country. The NORAD study was driven by the need for relevance and should meet that aim while being as rigorous as possible. It is the degree of rigour and lack of ability to study the evidence presented and come to the same conclusions that might detract from the validity of the report as an evaluation.

The Panel concluded that if such a study were to be repeated, it should consider three changes:

- First, to undertake a longer and more systematic analysis of the project documentation and a more in-depth and wide-ranging review of the relevant literature.
- Secondly, to have longer country visits in order to:
 - have more in-depth discussions with different stakeholders and with more beneficiaries, and to be able to scrutinise more carefully the differing and at times conflicting views on the impact of CSOs provided by different informed interlocutors and stakeholders;
 - pursue more vigorously the central issue of attribution; and
 - provide more documentation of why particular conclusions were drawn and judgements reached, especially where data and interviews produced differing assessments.
- Thirdly, based on the Panel's firm view that the wider impact of CSOs can only be assessed from the basis of an in-depth understanding of the context, culture and history of each country, to ensure that experienced evaluators and scholars from each country are included in any future Panel of assessors. (Riddell, 2013: 388).

In addition, the NORAD study was looking primarily at the impact of Norwegian-funded CSOs, but in reality looked more broadly. If it were possible to undertake a similar study that includes all of civil society, regardless of the donor, it would be more useful.

CASE STUDY 2: Reach, Relationships and Results: Case Studies of Australian NGOs' work in Education

Members of the ACFID Education Sector Working Group compiled 19 case studies from 11 Australian NGOs and undertook a narrative synthesis across the case studies to demonstrate that NGOs bring qualities and capacities to their work in education and development that are unique, efficient, strategic and often underestimated. The primary audience was DFAT (then AusAID), to inform ongoing policy dialogue between the Education Sector Working Group and AusAID education personnel, which began in 2012.

The case studies provide evidence that highlighted six key themes:

- NGOs can reach the unreached.
- NGOs work across the whole education sector but target the gaps.
- NGOs build long-term relationships at community level and strengthen civil society.
- NGOs work locally but have a range of relationships with government up to national level.
- NGOs are an effective, and often efficient, way of bringing about change at system level
- NGOs' work is technically sound, evidence-based and able to document long-term impact.

The ACFID Education Sector Working Group (now Community of Practice) was a proactive group. It had developed a good relationship with the education team at AusAID and this has continued into DFAT. The intent of the synthesis was always to demonstrate to DFAT the unique and necessary contribution of the NGOs to education. Both large and small ANGOs participated in providing case studies for the synthesis. The process, from conception to publication, took about a year in 2013, led by the two co-convenors of the Education Sector Working Group, who, with two others, formed a core group of four representatives from four NGOs.

The core group managed and implemented the whole process, with encouragement and assistance with final formatting and printing from ACFID and funds for publication (170 hard copies) provided by seven of the participating NGOs as well as a small ACFID contribution. Clear guidelines for case studies were developed, selectively seeking the key themes highlighted in the final report as follows:

'The compilation will highlight the impact of ANGO interventions, how they have engaged marginalised communities, have introduced innovations at the grassroots level and how their demonstrations of effectiveness have led to broader adoption and to changes in systems, policy and/or practice at higher levels.'

The guidelines also included an important note about language and evidence, emphasising that the purpose was to better explain the particular contribution of the NGOs to education and that this contribution can be significant:

'Case studies must be written from a programmatic perspective and avoid the language of marketing commonly associated with development "success stories". They should be critical and evidence-based. Case studies should demonstrate ANGO's distinctive contribution to education and how their interventions have influenced the broader education agenda in the countries.'

Coordination was a time-consuming process, firstly getting the case studies from the participating NGOs and then among the core team working out how to present the synthesis, which is in two parts – the synthesis itself followed by each of the case studies. The main

work of the narrative synthesis was undertaken by the two co-convenors. They distilled common experience from the case studies as relevant into the framework that had been worked out. The amount of additional work for the core group, especially the co-convenors, was considerable. Resourcing such projects is a big consideration.

Some of the case studies drew on material from evaluations, others did not. The quality of the case studies was generally good in terms of evidence, although some persisted in a level of agency branding in spite of the guidelines and one was less substantial in terms of content. It was deemed politically sensible to include all the case studies however, with the co-convenors deciding which material to use in the synthesis itself, providing a level of quality assurance.

The process was useful for the smaller agencies to see their case studies in print and 'endorsed' by ACFID and AusAID. It benefitted all the participating NGOs in the eyes of AusAID, who now had evidence that the NGOs had something serious to offer in the education sector.

Some of the case study material was used as the basis for a presentation session at the 'Development Futures' Conference in Sydney in November 2013. Once complete and published by ACFID, the core team also developed a two-page 'policy brief' for AusAID. Meetings were held with AusAID's education team and staff from post around the report, which was certainly visible to them and used at least internally within AusAID. The NGOs report being taken more seriously by AusAID's education team as a consequence. The synthesis may have had some influence on the 2014 design of the Basic Education Quality and Access in Lao PDR, which includes an NGO consortium component. As part of a broader advocacy and influence agenda, it is also felt the case studies (some of which were deliberately selected for their early childhood focus) had a significant influence on the incorporation of Early Childhood Care and Development as the first pillar in DFAT's most recent education strategy, developed in 2014-15. This marked a significant departure from previous education strategy iterations and was seen as a win for NGOs who had been individually and collectively advocating for the inclusion of early childhood care and development.

However, overall there was a sense of frustration that the document was not used to its full advantage in establishing a regular policy dialogue process with AusAID and that the amount of influence the report had with AusAID (now DFAT) is not really known. The level of influence may in part have been affected by the timing of the DFAT merger with AusAID. The project might have benefited from more thought going into planning for the strategic use of the document from the beginning of the process. Nonetheless the co-convenors felt it was worth the considerable effort.

CASE STUDY 3: Exploring the link between child and youth participation and development effectiveness

Interest in the role of child and youth participation and its influence on development effectiveness has been growing in recent years. In 2011 and 2012, following formal and informal discussions, members of the ACFID Child Rights Working Group identified this area as a gap in the research. Consultation held between interested Australian NGOs (ANGOs) led to the formation of a research and learning partnership to explore the issue through a dedicated research project.

The research project was a partnership between three ANGOs (Transform Aid International, ChildFund Australia, and Live & Learn), their in-country partner organisations and a university partner (the Institute for Sustainable Futures (ISF) at the University of Technology, Sydney), from 2013 to 2016. The research focussed on projects that involved children and youth in a participatory way. These included: youth-led environmental management and livelihood initiatives in Fiji; sport and child clubs in Laos, and child clubs in Nepal. The research was supported by an independent consultancy, InSIGHT Sustainability; and a Peer Review Group of child and development specialists.

The overarching objective of the research was to explore the links between child and youth participation and development effectiveness. A 'Development Effectiveness Analysis Mapping Tool' defined the most important aspects about development effectiveness as they relate to children and youth. These were:

- Participation in setting development priorities;
- Promotion of inclusion and equity, and reducing marginalisation;
- Knowledge sharing, mutual learning and collaboration;
- Personal and community development; and
- Socio-economic development.

The research sought to understand and learn more about this link, and build an evidence base to generate knowledge and inform learning to benefit in-country partners, ANGOs and the broader international development community. The three main research learning areas were:

- The contribution of child and youth participation to development effectiveness;
- The contribution of child and youth participation to in-country partner organisation outcomes and ANGO program outcomes; and
- Child and youth participation in research practice

A Joint Research and Learning Partnership was formalised between the participating ANGOs signed by the CEO/Director of each organisation. Contractual arrangements for the research were developed and signed between Transform Aid International, Live and Learn Fiji and ChildFund Australia (in the role of lead ANGO), as well with InSIGHT Sustainability and ISF for academic project support (e.g. help designing the research approach). A Research Plan was developed documenting roles and responsibilities, research aims and objectives, timelines and activities, field methods, research data collection and ethics. A particular focus was given to developing comprehensive Ethics Guidelines for the research.

The research was in two phases. The research was carried out through the joint efforts of the partners as follows (Phase 2):

In-country researchers in Fiji, Laos and Nepal: In-country partner staff led the research and a team of trained local researchers carried out the research in each of the nine research sites. In Laos and Nepal, researchers included youth. In-country researchers tested the

research tools and in some cases, adjusted them to fit the local context. In-country partners worked with communities to set up the research, collect and collate research participant responses and provided contextual background. In-country partner staff and researchers were trained by ANGO staff. They included a consultant who was familiar with the country and staff; and for Nepal Phase 2, an experienced academic and practitioner from the Peer Review Group volunteered who was familiar with Nepal.

ANGOs: ANGO staff were responsible for leading research training workshops in each country. They came up with the idea and first iteration of the Development Effectiveness Tool, co-designed the research, conducted quality reviews of data collected and co-wrote the Case Studies and Learning Paper.

ISF: ISF researchers co-designed the research approach and research tools with the ANGOs, and helped prepare the training workshops. They enabled a collaborative process with ANGO partners, and through them with in-country partners, to co-design all stages of the research. ISF analysed the research data and drafted the Learning Paper for research partner input.

Different research approaches were used for the three Learning Areas (phase 2), as each area had a different focus.

Learning Area 1 used a case study approach to explore the link between child and youth participation and development effectiveness. In this approach, children and youth described 'change stories' that happened because of their own participation. Illustrated Change Cards were used by children and youth to link their stories to the development effectiveness characteristics defined for this research.

Learning Area 2 used a self-assessment tool to explore the contribution of child and youth participation to program outcomes. In-country partners led the assessment to identify links between changes resulting from child and youth participation described by research participants in Learning Area 1 to see what extent they matched or contributed to their stated program objectives and in some cases their organisational objectives or vision.

Learning Area 3 used reflection questions to explore child and youth participation in research practice. Researcher daily debrief sheets, a post-research survey conducted using an online survey tool, and reflection/learning teleconferences with in-country partners were used by the research partners to collect this information.

Research data was collected for the first phase in 2014. Analysis was conducted in three tiers: first by in-country partners; then by ANGOs; and finally by ISF in synthesising the data and preparing the Mid Term Learning Paper (2015). The second phase built on the learnings from the first phase and was undertaken in 2016. Data was collated by in-country partners then reviewed by ANGOs to ensure all relevant data was captured before ISF synthesised the data and provided responses to the research questions, leading to a final Learning Paper in 2017 that was being finalised at the time of preparing this case study. Shorter products will also be developed for advocacy purposes.

Main findings from the research

The research provided evidence that shows how child and youth participation can make valuable contributions to development effectiveness, for example:

1. Child and youth participation led to positive changes within children and youth themselves (e.g. improved interpersonal skills and self-confidence), these changes then enabled children and youth to influence broader change within families and communities.
2. Child and youth participation had benefits to personal and community development such as improved equity, social inclusion and self-esteem among children and youth. These contributed to longer term benefits, such as socio-economic development, good health and education.

The partners plan to promote the use of the evidence and findings of this research by development practitioners in two main ways – in advocacy about the value of child and youth participation in development programs and to consider including the contribution of children and youth and the tools used in the research, such as the development effectiveness framework and the change story methodology in project monitoring and evaluation

- Technical challenges included the longer than anticipated time it took to develop the tools for Phase 2, including the need to make adjustments to match the in-country context. While the methodology provided rigour in the findings, there were issues faced in the analysis of the data when the first learning paper did not capture its richness and there were also difficulties in drafting the second learning paper.
- In terms of localising the approach, the partners reflected that the approach was a bit 'top down' and they should have involved in-country staff earlier and more substantially in the process as well as involving children and youth. Funding was insufficient for a more collaborative process. Partner training for data collection and collation was part of Phase 2 as well as the involvement of youth in the activity.
- The activity over-reached on the resources available. Adequate human resourcing is essential, with all partners needing a key person identified as the driver of the research. Additionally, both a budget and a commitment to contribute to not only each organisation's own data collection activities, but to the wider ANGO budget requirements, is needed. ISF's funding model was expensive.
- Collaboration and the commitment of the partners was seen as a strong point, but obtaining buy-in from each ANGO (including management, programs teams and other departments) was essential.
- Conducting research in partnership, including with a University partner, required technical support be provided (e.g. through technical advice, provision of research training or upskilling) in such a way that incorporated academic rigour whilst also addressing the needs of the ANGOs and its partners.
- The partnership needed regular face to face meetings amongst partners. The University partner was not always invited as they were on a contract with limited hours assigned. The time required for ANGO partners was considerable and difficult to devote when this activity was considered a minor part of their job. Nonetheless, partners appreciated the significant learning activity provided by the activity and are applying some of this learning to their own monitoring and evaluation systems.

Entering into partnership with a local University and the possible inclusion of stakeholders that the activity is designed to influence throughout the activity were two suggested ways in which some of these issues might be addressed.

CASE STUDY 4: Australia Africa Community Engagement Scheme

The Australia Africa Community Engagement Scheme (AACES) was a partnership between DFAT, ten Australian NGOs and their in-country partners across 11 Sub-Saharan African countries. The ten Australian NGOs (ActionAid, The Australian Foundation for the Peoples of Asia and Pacific Ltd, Anglican Overseas Aid, CARE Australia, Caritas Australia, Marie Stopes International Australia, Oxfam Australia, Plan International Australia, WaterAid Australia and World Vision Australia) delivered community-based interventions over a five year period (2011 – 2016), across the three sectors of water and sanitation, food security and maternal and child health. The community based interventions targeted the most marginalised groups with a particular focus on women, children, people with disability and people vulnerable to disaster. The design of individual NGO activities and their monitoring and evaluation was part of the program.

AACES had the following objectives:

Objective 1: Marginalised people have sustainable access to the services they require.

Objective 2: Development programs, including AACES, are strengthened, particularly in their ability to target and serve the needs of marginalised people through learning, collaboration and exchange among AACES NGOs and between AACES NGOs and DFAT.⁷

Objective 3: To better inform the Australian public about development issues in Africa (retired from July 2014).

The monitoring and evaluation framework for AACES was positioned within the partnership approach for the whole program. It had two levels of enquiry:

- Monitoring and evaluation within each of the NGO projects;
- Monitoring and evaluation for the overall AACES program.

This case study is focused on the second level, but recognises it is intrinsically linked to the approach used for the first level. Project level NGO monitoring and evaluation provided the bulk of the information used to assess and report on the progress against the AACES objectives. NGOs were encouraged to use their existing systems for performance management, but to ensure these met a series of quality requirements:

- Assessing not just the quality of activities, but also the positive and negative outcomes that resulted;
- Assessing different levels of involvement, and outcomes for women and men as well as for people with disability;
- Using a mix of methods, and gathering information from a range of sources;
- Collecting baseline information for a small selection of outcome areas and maintaining flexibility within the monitoring and evaluation approach to accommodate the likely changes in project design.

An external consultant was engaged to facilitate and support DFAT and the NGOs to agree on outcome statements for the program objectives. This significant step allowed DFAT and

⁷ This objective was modified after the midterm review. The original objective 2 read 'AusAID policies and programs in Africa are strengthened particularly in their ability to target and serve the needs of marginalised people'.

the NGOs reach a common understanding on what success looked like and establish boundaries around what would be judged. A reporting template agreed up front in the project design was also important in ensuring the accountability needs were defined at the start of the program, as well as allowing for flexible reporting in an annex where NGOs could describe the complexity of the work and include community stories.

The consultant worked closely with DFAT and the NGOs to identify a set of quantitative indicators which all NGOs could report against, for example the number of additional people with access to sustainable and safe water, number of farming households adopting new technologies, and number of children that receive age appropriate immunisation. These indicators, disaggregated by gender, disability and other categories, enabled the NGOs to demonstrate the quantitative value of their work and allowed DFAT to demonstrate the reach and inclusion of the program in their public accountability reporting.

The NGOs were encouraged to develop theories of change (as opposed to logical frameworks) for their particular projects with a view to contributing towards the agreed outcomes. These theories, developed in partnership with the NGOs' local partners in country, were identified as a strength by those partners:

'We like having the time to develop the M&E. In the past for this organisation and for partners it has tended to be a quantitative process, it has been good to have a process with higher-level outcomes and considerations and have time to work with partners to develop this and get them up to speed.'

'The theory of change approach is now being used in other projects in the national offices'.⁸

These theories of change formed the basis of the NGOs' project-level monitoring and evaluation. A theory of change for the whole program was developed at the time of the mid-term review.

Towards the beginning of the implementation phase, a Program Steering Committee (PSC, comprising 12 people, one from each Australian NGO plus two representatives from DFAT) was convened to oversee the program. The leadership of the committee moved from DFAT to the Australian NGOs to the in-country partners. The last two years of the program saw in-country partner NGOs represented in the PSC as Vice-Chair and Chair. AACES included a Resources Facility, based in Nairobi to support program management tasks and responsibilities, including coordination, communication and program level monitoring.

Program level monitoring and evaluation took a very participatory approach. After the first year of implementation, and as flagged in the design, a consultant was engaged to facilitate a monitoring and evaluation systems review. The review began with Australian NGOs, then with in-country partners in three countries and culminated in a workshop in Africa with over 60 participants. While there were many practical benefits from the review in terms of improvements to NGO monitoring and evaluation systems, perhaps the most significant benefit was in creating a culture of trust and enquiry between the NGOs that continued throughout the life of the program. The key features of the program-level monitoring included:

1. **Annual Review Workshops** – these were the most valued aspect of program-level monitoring, evaluation and learning. Preceded by field visits to project sites and project level review and planning workshops. The location of the annual workshops varied between the target countries and were attended by DFAT, Australian NGOs, in-Africa partners, and occasionally beneficiaries and recipients were invited. The

⁸ Source: Overview and Lessons from the AACES Design Process, Kelly, L and Kenway, J. 2012

workshops helped NGO learning, building capacity of the partners, addressing management issues jointly and strengthening the trust and partnership between the NGOs and DFAT as well as among the NGOs themselves. They were an opportunity for NGOs to review lessons, achievements, and effectiveness while confirming messages and themes for the Annual Program Report.

2. **Annual Program Report** – This was a synthesis of NGO narrative reports (NGOs provided six monthly reports or updates and annual narrative reports assessing progress against the program objectives and outcome areas.) It was prepared by the Resource Facility to identify key messages and trends and illustrate individual highlights by assembling a range of case studies. It was complimented with additional evaluation and research studies as determined by the annual report working group selected by the PSC.
3. **Peer Review** – The NGOs peer reviewed each other’s designs and draft monitoring and evaluation approaches in Australia and in several countries in Africa through the first 10 months of the design process. Other examples of peer review included the PSC reviewing the effectiveness of the resource facility each year, and the NGOs reviewed the quality of their partnership each six months.
4. **Mid-term and Final Reviews** - NGO reports provided the main data for the reviews; however the external review teams collected additional data. An external review of the effectiveness of the management approaches was also carried out during the midterm and final reviews for the AACES program.
5. **Value for Money Assessment** - NGOs were explicitly asked to focus on value-for-money within their individual monitoring, evaluation and learning frameworks. Each NGO partner shared their organisational approach to value-for-money and the program developed an overarching framework through a lengthy, consultative process. This was reported across the program as in the Annual Program Report.

Figure 1. Summary of AACES monitoring and evaluation approach⁹

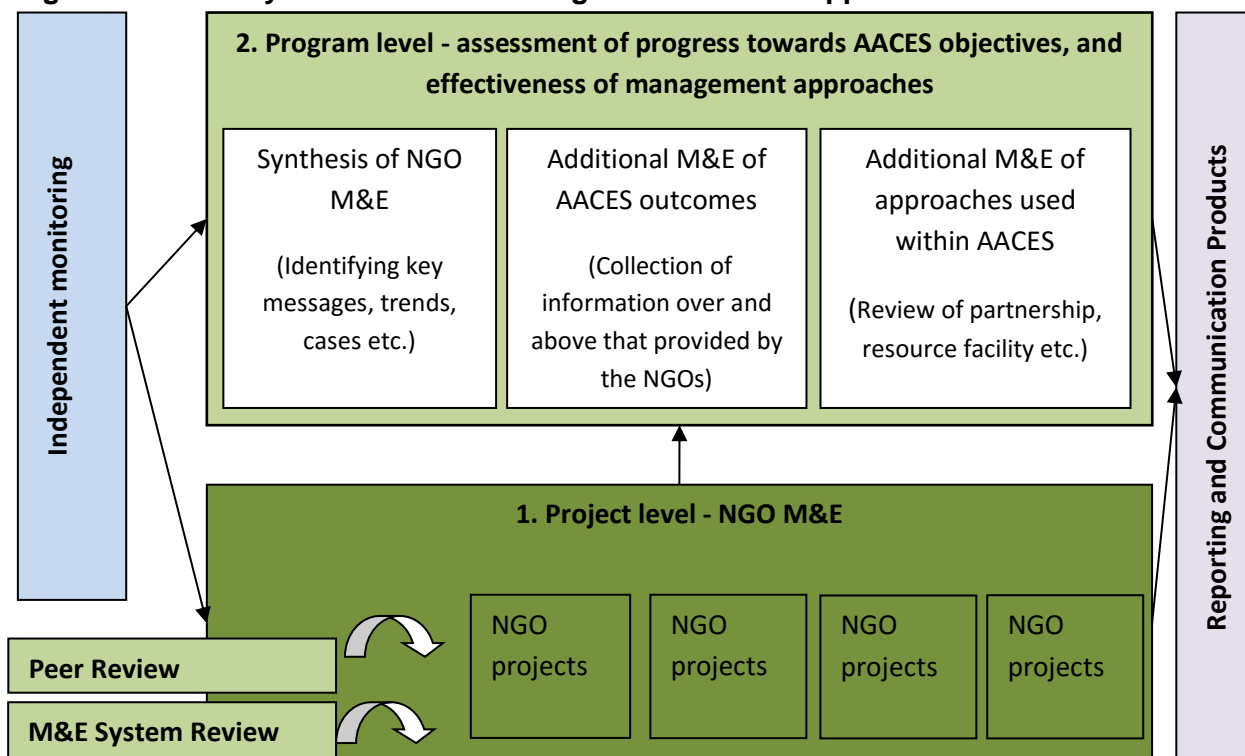


Figure 1 illustrates the overall approach to monitoring and evaluation. In practice, the key features of annual reporting were the aggregated common indicators and individual case studies. Common cross-cutting themes were extracted from the case studies to broadly illustrate some of the commonalities, for example in mobilisation and organisation of men and women in communities, facilitating service delivery to poor and marginalised people, enhancing social inclusion and gender equality, capacity development and social accountability. To some degree this demonstrated the unique contribution of the community based NGO approach, but coherently communicating the breadth and diversity of this program remained a challenge.

The monitoring, evaluation and learning approach adopted by the AACES was highly valued by all the partners and its design received an award for excellence from the Australian Evaluation Society in 2016. It was distinctive in a number of ways:

Localised approach

The monitoring and evaluation was heavily centred close to the action in Africa and included the in-Africa NGO partners. The annual workshops to reflect on the program, allowed in-Africa partners to share in the review, benefit from capacity building and participate in the management of the program. At a country level, the partnership approach helped local NGOs to collectively advocate with the national governments, for example, around water programming in Tanzania and budgeting for water in Malawi. The Malawi AACES Network NGOs shared tools for monitoring, such as community digital stories. Following completion of the program, the Malawi AACES Network still appears to be active. The longevity of this network was helped by the common sector of projects in that country as well as the locations of the local NGOs. In other countries, the variety of sectors and remote locations of some of the NGOs was less conducive to a sustained network.

⁹ Source: AACES Program Review 2011-16, DFAT 2016, page 27

Accountability and learning

The approach met both the accountability needs of DFAT and the learning needs of the activities themselves as activities were adjusted on the basis of learning. The deliberate building of trust among the partners and the shared power relationship between donor and NGO partners were both key elements to allow the heavy learning focus. One aspect of this was having champions of the approach among the DFAT staff. It is a testament to the commitment of the DFAT staff and their strong partnership with the AACES NGOs that the program survived major changes to Australian Government policy in 2013-14 including budget restructuring and a merger of AusAID with DFAT.

Resourcing

Monitoring and evaluation was identified in the initial design as an important aspect of this program. NGOs were encouraged to use up to 10% of their budget for monitoring, evaluation and research activities, and up to 10% of the overall AACES budget had similarly been allocated towards management, monitoring, evaluation and research. The development of monitoring and evaluation approaches in the NGOs and with partners, the annual reviews and meetings, engagement in the PSC, reporting on projects and across the program required a considerable investment of NGO and DFAT staff time, including Canberra-based staff and a full time position based in Nairobi as well as consulting costs. Allowances for these activities were provided in the budget.

Embracing complexity

The approach to monitoring and evaluation did not attempt to oversimplify the complexity of the development challenges being faced in the eleven countries nor the complexity of the program itself. The non-prescriptive approach to monitoring and evaluation systems was key as well as the development of agreed outcome statements. Trying to find coherence in monitoring and evaluating a disparate program across multiple sectors, multiple countries and multiple agencies was a major challenge, with the added complexity of being driven from the ground up, in other words the design was not prescribed. A focus on cross-cutting issues such as gender and disability was a helpful strategy for managing aspects of the complexity. The program configuration was an artificial construct created by the DFAT funding window; nothing else bound together these countries, sectors or agencies. A focus on monitoring and evaluation at the level of a single country (rather than eleven countries across the continent of Africa) might be another helpful strategy in future applications.

CASE STUDY 5: OXFAM Australia: Strategic Plan (2014-19) Reporting

Brief history of reporting approach

The 2015 Outcomes Report was the first assessment of OXFAM Australia's progress towards achieving their 2014-19 Strategic Plan goals. The Outcomes Report focused on three out of six of the Strategic Plan goals and five results. It provided an analysis of progress towards the outcomes that underpin each result and an assessment of the effectiveness of the strategies used to achieve these outcomes. The Outcomes Report was designed to be used by the OXFAM Australia Board and Management Team to monitor the overall performance of OXFAM Australia programs and to contribute to thematic strategy and program learning and development.

The Outcomes Report was based on a narrative analysis of program evaluations (mid- and end evaluations), annual reports of OXFAM Australia, and interviews with program staff. External support was engaged to analyse program documentation against the prioritised questions and indicators and provide an external perspective on the strength of the evidence for program contributions to Strategic Plan results. There was also a process of review by program staff, thematic advisors and senior management. This was another layer of both analysis and integrating dissemination into the process.

The Outcomes Report provided a view across the OXFAM Australia portfolio of projects that had previously not been available and provided insights that corrected assumptions about the depth and breadth of OXFAM Australia's program investments. It also highlighted gaps in thematic and program strategies and risks. These have been addressed via subsequent initiatives. However, the Outcomes Report alone was not able to fulfil all the functions that OXFAM Australia needed. Learning from the production of OXFAM Australia's 2015 Outcomes Report included the need to:

- Clarify and gain agreement among key audiences on the purpose and uses of Strategic Plan Reporting early in the process
- Invest in ensuring effective dissemination and use of Strategic Plan Reporting and thematic learning products

As a result a new and diversified reporting system was developed for 2016-17, based around a suite of products. The new approach is more informed by a learning approach. Individual products within the suite are structured so that they address strategic learning questions and provide analysis at specific points in the OXFAM business development cycle to contribute to program development and decision making about the portfolio.

Current OXFAM Australia Reporting System

OXFAM Australia's reporting system integrates with that of OXFAM International, but draws a boundary around OXFAM Australia's portfolio. OXFAM Australia's Strategic Plan contains six goals, each with multiple 'results', for example:

Goal 2 Gender justice	Result 2.1 More women and women's organisations will have decision-making roles and influence at local and national levels.
	Result 2.2 More women and girls will be living free of violence and the fear of violence.

The aim is to report every two years against the Strategic Plan through a Strategic Plan Reporting Summary Assessment, drawing on a suite of products:

1. OXFAM Australia's Portfolio Analysis

Data (geography, implementation dates, budget, expenditure, outcomes and beneficiary numbers) is collated from across all initiatives to provide a high level overview of the program portfolio every six months. Over time this enables the analysis of investments by goal. It draws on data from all sources, including OXFAM International Output Reporting which seeks disaggregated beneficiary numbers for each OXFAM project or Program and against Strategic Plan goals.

2. OXFAM Australia Evaluation Summary

This lists planned and completed evaluations with links to evaluation reports.

3. OXFAM Australia Evaluation Reviews

Desk-based narrative synthesis of program evaluations, to summarise an analysis of outcomes and strategies against strategic plan goals and themes. Also assesses evaluation quality and tracks implementation of recommendations. This product is prepared every six months to facilitate the use of learning from program evaluations and identify issues and trends across programs.

4. OXFAM Australia Program Performance Reviews

These annual reviews provide an overview by analysing a sample of ongoing program and project strengths and challenges against agreed criteria and progress towards anticipated outcomes, including highlighting challenges to achieving outcomes.

5. OXFAM Thematic Frameworks

Conceptual thematic frameworks will establish the high level outcomes and strategies (like theories of change) to achieve those outcomes across thematic areas. These frameworks will be used to explain OXFAM's development approach in external communications and also provide the framework for assessing thematic strategies and outcomes.

6. OXFAM Australia Strategic Outcomes Analysis

This is an in-depth analysis, using narrative synthesis of evaluations, of programs and projects within a thematic area, to identify factors that contribute to outcomes and effectiveness of strategies against one or two prioritised strategic plan goals and results.

OXFAM Australia's reporting products are complemented by OXFAM International products including OXFAM International Output Reporting and OXFAM International Outcome Areas Evaluations. OXFAM International products also include a third party NGO Partner Survey (see Box 3).

Box 3: NGO Partner Survey

An NGO partner survey tool has been used since 2010 by a group of 25 NGOs based in Europe and the USA to obtain anonymous feedback from their partner organisations on their performance. OXFAM makes use of the survey as part of its international reporting. It is an interesting example of a tool that specifically addresses the issue of downward accountability. This is relevant for those Australian NGOs that work primarily through local partner organisations, as opposed to those that implement directly.

An independent organisation, Keystone, surveys the southern partners of the NGOs and reports back to the NGOs. The latest available public report is in 2011. Partners were asked to rate and comment on different aspects of the NGOs' performance, with a guarantee that the NGOs would not be able to identify who said what about them. 1,067 out of 2,733

partners responded (39%). According to Keystone's website, to date, 77 NGOs have used the survey (including Plan International, OXFAM International and Save the Children, UK, US and International) and 18 have repeated the process (Keystone Accountability, 2016).

The findings are interesting, showing significance variations in performance across the NGOs. The southern partners particularly value the strong relationships and understanding of contexts by the NGOs. 'Respondents say that most northern NGOs provide them with reporting formats and that the monitoring they do for northern NGOs helps them improve what they do' (Keystone Accountability, 2011). The results showed that the southern partners have an average of five NGO or donor partners. Consistently lower ratings were around the limited flexibility to make changes, funding limitations, and limited engagement on strategy.

Oxfam's international approach to assessing outcomes against the Oxfam International Strategic Plan (OSP) goals involves evaluations of a sample of programs from across the confederation. While the OSP evaluations will assess broad programming trends and effectiveness, it is unlikely they will provide a detailed analysis of outcomes given the breadth of the Oxfam portfolio. It is also likely that only a few OXFAM Australia supported programs will be included in the sample for the OSP evaluations.

A common analysis framework for the suite of products includes the Strategic Plans goals and results and a set of strategic evaluative questions:

Outcomes

- What positive and negative changes have OXFAM Australia supported programs contributed to against prioritised Strategic Plan goals and results?
- Who has benefited and how have they benefited? What benefits have marginalised groups experienced?
- Are the changes at local, regional, national or global levels?

Effectiveness

- What strategies are effective/not as effective in enabling change at different levels in programs and projects?
- What strategies and tactics have worked in supporting people, communities and civil society to influence decision makers? How effectively have marginalised people been supported to take leadership roles and influence decision makers?
- How effectively has OXFAM worked with implementing and strategic partners?
- What has been OXFAM's contribution to program/project outcomes and effectiveness?

Discussion

The OXFAM Australia approach will continue to emerge and develop. It takes a learning approach, with continuous learning loops, but there is still work to be done to align the individual pieces of analysis to their use. It is a big organisation with a dispersed focus, making it difficult to find the right moment when particular reporting will help inform particular programming decisions in the business cycle, such as budgets, country engagement strategies, project designs and monitoring, learning and evaluation approaches or the development of thematic strategies. Timing is one challenge but also the specific messaging of the products is another. Each of the products needs to be fit for purpose and available at the right moment to the right audience and this requires identifying the right connections within OXFAM Australia's business processes.

OXFAM's performance system is aligned directly to OXFAM's Strategic Plan and should assist OXFAM in performing its mission and vision and directly help with advocacy efforts for example. The system is designed to embrace complexity and not be reductionist. The suite of products should ultimately support reflection on whether Oxfam's contribution addresses

root causes, will support positive change in people's lives and enables structural and systemic change so that communities and their organisations can influence decisions that affect them. OXFAM Australia is working towards an environment where they will be more focussed on strategic contributions to the development and implementation of OXFAM country programs in prioritised geographic and thematic areas. In this context, OXFAM Australia will need information and analysis that supports decisions about financial, strategic and technical investments. This will contribute to internal decision-making, learning and program and thematic strategy development and also support engagement with institutional donors, supporters and the public. Many products in the Strategic Plan Reporting suite are also designed and used as OXFAM resources. For example, the Violence Against Women and Girls Monitoring, Evaluation and Evaluation Guidance will be an Oxfam resource.

In accountability terms the system will meet DFAT's requirements. Accountability to partners and communities is more complex, because OXFAM International now directly manages programs in-country, although OXFAM Australia is involved in various ways: country programs are managed by a Regional Director and OXFAM Australia as an OXFAM Affiliate is a member of regional and country governance groups. OXFAM Australia is involved in the OXFAM International Monitoring, Learning and Evaluation Governance Group and several other Oxfam International Monitoring, Learning and Evaluation initiatives. As an affiliate OXFAM Australia also support country Monitoring, Evaluation and Learning including for contracts that are managed by Oxfam Australia. However, downward accountability is variable, communities tend to be providers of data. Evaluation results are shared however and publicised and local partners are involved through their MEL processes. In addition, OXFAM International draws on a third party partner survey (see Box 3).

It is quite a complex system, and as yet untested in practice. Two OXFAM Australia staff are assigned to its implementation, one for the portfolio analysis and the other (who has other responsibilities) working closely with internal and external experts to implement the remainder. Maintaining this level of staffing and sufficient budget to implement will be important. One of the difficulties is maintaining a system of this nature is ensuring consistency of expectations and commitment from the Board or OXFAM Australia senior management level over the life of the Strategic Plan.

Another risk (which arose during the 2015 Outcomes Report) relates to the quality of program documentation. Some programs had good documented evidence of outcomes and program effectiveness and others did not. Program evaluations are not consistently focused on assessing outcomes and the level of rigour in testing the effectiveness of program strategies is variable. Program evaluations often do not specifically assess how OXFAM Australia contributed to outcomes

A further risk is whether this approach provides a coherent (enough) view across the portfolio and can support decisions at different levels and at the right time; how to continue to develop the relationships between supply and demand so that the analysis pieces in the Strategic Plan Reporting suite are responding to priority questions among key audiences.

References

- ABUOM, A., ALEMU, K., EVENSMO, I., HAFILD, E., KRUSE, S.-E. & RIDDELL, R. 2012. Tracking Impact: An exploratory study of the wider effects of Norwegian civil society support to countries in the South. Oslo: Norwegian Agency for Development Cooperation.
- ACFID. Stakeholder Workshop Demonstrating outcomes and impact across different scales, 4 May 2017 Melbourne.
- ACFID 2015. Developing an Effectiveness Framework: A toolkit for small and medium sized NGOs. *In*: ACFID (ed.). Canberra.
- ACFID 2016. Principles and Guidelines for ethical research and evaluation in development. Canberra: ACFID.
- ACHARYA, A., DE LIMA, A. T. F. & MOORE, M. 2006. Proliferation and fragmentation: Transactions costs and the value of aid. *The journal of development studies*, 42, 1-21.
- ANDREWS, A. 2014. Downward accountability in unequal alliances: explaining NGO responses to Zapatista demands. *World Development*, 54, 99-113.
- AUSTRALIAN EVALUATION SOCIETY 2013. Guidelines for the Ethical Conduct of Evaluations. Australian Evaluation Society.
- BANKS, N., HULME, D. & EDWARDS, M. 2015. NGOs, States, and Donors Revisited: Still Too Close for Comfort? *World Development*, 66, 707-718.
- BELL, S. A. & AGGLETON, P. 2012. Integrating ethnographic principles in NGO monitoring and impact evaluation. *Journal of International Development*, 24, 795-807.
- BETTER EVALUATION. Available: <http://www.betterevaluation.org/> [Accessed March 2017].
- BETTER EVALUATION. *Synthesise Data Across Evaluations* [Online]. Available: http://www.betterevaluation.org/en/plan/synthesize_value/synthesize_across_evaluations [Accessed March 2017].
- BOND. *Impact Builder* [Online]. London: BOND. Available: <https://www.bond.org.uk/resources/impact-builder> [Accessed March 2017].
- CAPLAN, K. & JONES, D. 2002. Measuring the effectiveness of multi-sector approaches to service provision. *Water Science and Technology: Water Supply*, 2, 189-196.
- CENTER FOR GLOBAL PROSPERITY 2013. The Index of Global Philanthropy and Remittances. Washington: The Hudson Institute.
- CHRISTENSEN, D. & WEINSTEIN, J. M. 2013. Defunding dissent: restrictions on aid to NGOs. *Journal of Democracy*, 24, 77-91.
- CHRISTIAN AID. *Promoting behaviour change through faith leaders in Burundi* [Online]. Available: <http://programme.christianaid.org.uk/programme-policy-practice/case-study/promoting-behaviour-change-through-faith-leaders-burundi>.
- DEGNBOL-MARTINUSSEN, J. 2003. *Aid : understanding international development cooperation*, New York, Zed Books.
- DEVELOPMENT ASSISTANCE COMMITTEE 1991. Principles for Evaluation of development assistance. Paris: Organisation for Economic Cooperation and Development.
- EYBEN, R. 2013. Struggles in Paris: The DAC and the Purposes of Development Aid. *European Journal of Development Research*, 25, 78-91.
- EYBEN, R. 2015. Uncovering the politics of evidence and results. *Politics of results and evidence in international development: Playing the Game to Change the Rules*, 19-38.
- FLYVBJERG, B. 2006. Five misunderstandings about case-study research. *Qualitative inquiry*, 12, 219-245.
- FROT, E. & SANTISO, J. 2009. Crushed aid: fragmentation in sectoral aid.
- GORE, C. 2013. The new development cooperation landscape: actors, approaches, architecture. *Journal of international development*, 25, 769-786.
- GREEN, D. 2015. Fit for the Future? Development Trends and the role of international NGOs. *Oxfam Discussion Papers*. Oxford, UK: OXFAM.
- GREENHILL, R. & PRIZZON, A. 2012. Who foots the bill after 2015? What new trends in development finance mean for the post-MDGs. *London: Overseas Development Institute (Working Paper 360)*.
- HELLIKER, K. 2007. Marx, Weber and NGOs. *South African Review of Sociology*, 38, 120-133.
- HOLMA, K. & KONTINEN, T. 2011. Realistic evaluation as an avenue to learning for development NGOs. *Evaluation*, 17, 181-192.
- HOLVOET, N. & ROMBOUTS, H. 2008. The challenge of monitoring and evaluation under the new aid modalities: experiences from Rwanda. *The Journal of Modern African Studies*, 46, 577-602.

- HOLZAPFEL, S. 2014. The role of indicators in development cooperation: an overview study with a special focus on the use of key and standard indicators.
- HORTON, K. 2010. An Appeal to Aid Specialists. *Development Policy Review*, 28, 27-42.
- HOWELL, J. & HALL, J. 2012. Working Beyond Government: Evaluation of AusAID's Engagement with Civil Society in Developing Countries. Canberra: Australian Agency for International Development.
- ITAD 2014. Investing in monitoring, evaluation and learning: Issues for NGOs to consider. UK: Big Lottery Fund, Bond, Comic Relief, NIDOS and DFID.
- KELLY, L. & CHAPMAN, R. 2010. A Process to Define and Enhance NGO Effectiveness. Canberra: ACFID.
- KEYSTONE ACCOUNTABILITY 2011. Keystone Performance Surveys: NGO Partner Survey 2010 Executive Summary. London: Keystone Accountability.
- KEYSTONE ACCOUNTABILITY. 2016. *International Non-Governmental Organization Survey* [Online]. London. Available: <http://keystoneaccountability.org/international-non-governmental-organization-survey/> [Accessed March 2017].
- KILBY, P. 2010. *NGOs in India : The challenges of women's empowerment and accountability*, London, Taylor & Francis.
- KIM, E. M. & LEE, J. E. 2013. Busan and Beyond: South Korea and the transition from aid effectiveness to development effectiveness. *Journal of International Development*, 25, 787-801.
- KINDORNAY, S. 2011. From Aid to Development Effectiveness: A Working Paper. Canada: The North-South Institute.
- LEFTWICH, A. 2011. Thinking and Working Politically: What does it mean, why is it important and how do you do it? in Politics, Leadership and Coalitions in Development: Policy Implications of the DLP Research Evidence, Background Paper. *Research and Policy Workshop*. Frankfurt, Germany: The Developmental Leadership Program.
- LISSNER, J. 1977. *The Politics of Altruism: a study of the political behaviour of voluntary development agencies*, Lutheran World Federation, Department of Studies.
- MAWDSLEY, E., SAVAGE, L. & KIM, S. M. 2014. A 'post-aid world'? Paradigm shift in foreign aid and development cooperation at the 2011 Busan High Level Forum. *The Geographical Journal*, 180, 27-38.
- MAYNE, J. 2001. Addressing attribution through contribution analysis: using performance measures sensibly. *The Canadian Journal of Program Evaluation*, 16, 1.
- MAYNE, J. 2004. Reporting on outcomes: Setting performance expectations and telling performance stories. *The Canadian Journal of Program Evaluation*, 19, 31.
- MAYNE, J. 2009. Building an Evaluative Culture: the key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24, 1.
- MAYOUX, L. & CHAMBERS, R. 2005. Reversing the paradigm: quantification, participatory methods and pro-poor impact assessment. *Journal of International Development*, 17, 271.
- MENOCAL, A. R. 2014. Getting real about politics: From thinking politically to working differently. Overseas Development Institute.
- MOLENAERS, N., JACOBS, B. & DELLEPIANE, S. 2014. NGOs and aid fragmentation: the Belgian case. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 25, 378-404.
- NATSIOS, A. 2010. The Clash of the Counter-bureaucracy and Development.
- NGO ADVISOR. *Top 100 NGOs* [Online]. Geneva. Available: <https://www.ngoadvisor.net/top100ngos/>.
- NILSSON, D., BAXTER, G., BUTLER, J. R. & MCALPINE, C. A. 2016. How do community-based conservation programs in developing countries change human behaviour? A realist synthesis. *Biological Conservation*, 200, 93-103.
- O'DWYER, B. & UNERMAN, J. 2008. The paradox of greater NGO accountability: A case study of Amnesty Ireland. *Accounting, Organizations and Society*, 33, 801-824.
- OECD 2007. Outcome Document from Workshop on Development Effectiveness in Practice: Applying the Paris Declaration to Advancing Gender Equality, *Environmental Sustainability and Human Rights*. Dublin, Ireland.
- OECD. 2011. *The Busan Partnership for Effective Development Co-operation* [Online]. Available: <http://www.oecd.org/dac/effectiveness/busanpartnership.htm> [Accessed 14 August 2015].
- OECD/DAC. *DAC Criteria for Evaluating Development Assistance* [Online]. Available: <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.
- OFFICE OF DEVELOPMENT EFFECTIVENESS 2014a. Learning from Australian Aid Operational Evaluations Canberra: DFAT.
- OFFICE OF DEVELOPMENT EFFECTIVENESS 2014b. Lessons from Australian Aid: 2013 report on independent evaluation and quality assurance. Canberra: Department of Foreign Affairs and Trade.
- OPEN FORUM FOR CSO DEVELOPMENT EFFECTIVENESS. The Siem Reap CSO Consensus on the International Framework for CSO Development Effectiveness. The Second Global Assembly Open Forum for CSO Development Effectiveness, 2011 Cambodia. Open Forum for CSO Development Effectiveness.

- PAWSON, R. 2006. *Evidence-based policy: a realist perspective*, Thousand Oaks, SAGE.
- PAWSON, R. 2014. *The Science of Evaluation: A Realist Manifesto*, Thousand Oaks, Sage.
- PAWSON, R. & TILLEY, N. 1997. *Realistic evaluation*, Thousand Oaks, Sage.
- RAMALINGAM, B. 2013. *Aid on the Edge of Chaos*, Oxford University Press, USA.
- REIMANN, K. D. 2006. A View from the Top: International Politics, Norms and the Worldwide Growth of NGOs. *International Studies Quarterly*, 50, 45-67.
- RIDDELL, R. 2007. *Does foreign aid really work?*, Oxford; New York, Oxford University Press.
- RIDDELL, R. C. 2013. Assessing the Overall Impact of Civil Society on Development at the Country Level: An Exploratory Approach. *Development Policy Review*, 31, 371-396.
- ROCHE, C. 2009. Oxfam Australia's experience of 'bottom-up' accountability. *Development in practice*, 19, 1009-1022.
- ROCHE, C. & KELLY, L. 2012. The Evaluation of Politics and the Politics of Evaluation. Developmental Leadership Program.
- ROGERS, P. J. 2008. Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14, 29-48.
- SCHMITZ, H. P., RAGGO, P. & BRUNO-VAN VIJFEIJKEN, T. 2012. Accountability of transnational NGOs: Aspirations vs. practice. *Nonprofit and Voluntary Sector Quarterly*, 41, 1175-1194.
- SCRIVEN, M. 1994. Evaluation as a discipline. *Studies in Educational Evaluation*, 20, 147-166.
- SIMISTER, N. 2016. Summarising portfolio change: results frameworks at organisational level. *M&E Paper 10* [Online].
- SNILSTVEIT, B., OLIVER, S. & VOJTKOVA, M. 2012. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness*, 4, 409-429.
- SNILSTVEIT, B., STEVENSON, J, MENON, R., PHILLIPS, D., GALLAGHER, E, GELEEN, M, & JOBSE, H., SCHMIDT, T AND JIMENEZ, E 2016. The impact of education programmes on learning and school participation in low- and middle-income countries: a systematic review summary report. London: International Initiative for Impact Evaluation (3ie).
- THE WORLD BANK 2016. Evaluating the Evidence for Historical Interventions Having Reduced HIV Incidence: A Retrospective Programmatic Mapping Modelling Analysis Washington DC: The World Bank.
- VÄHÄMÄKI, J., SCHMIDT, M. & MOLANDER, J. 2011. Review: results based management in development cooperation. *Riksbankens Jubileumsfond, Stockholm*.
- WILLETTS, J. & CRAWFORD, P. 2007. The most significant lessons about the Most Significant Change technique. *Development in Practice*, 17, 367-379.
- WORLD VISION 2014. Compendium of Indicators for Measuring Child Well-being Outcomes. World Vision International.

ACFID Member List

As at January 2018

Full Members:

- ACC International Relief
- Act for Peace - NCCA
- ActionAid Australia
- Action on Poverty
- Adara Development Australia
- ADRA Australia
- Afghan Australian Development Organisation
- Anglican Aid
- Anglican Board of Mission - Australia Limited
- Anglican Overseas Aid
- Anglican Relief and Development Fund Australia
- Asia Pacific Journalism Centre
- Asian Aid Organisation
- Assisi Aid Projects
- Australasian Society for HIV, Viral Hepatitis and Sexual Health Medicine
- Australia for UNHCR
- Australia Hope International Inc.
- Australian Business Volunteers
- Australian Doctors for Africa
- Australian Doctors International
- Australian Himalayan Foundation
- Australian Lutheran World Service
- Australian Marist Solidarity Ltd
- Australian Medical Aid Foundation
- Australian Mercy
- Australian Red Cross
- Australian Respiratory Council
- AVI
- Beyond the Orphanage
- Birthing Kit Foundation (Australia)
- Brien Holden Vision Institute Foundation
- Bright Futures Child Aid and Development Fund (Australia)
- Burnet Institute
- Business for Millennium Development
- CARE Australia
- Caritas Australia
- CBM Australia
- ChildFund Australia
- CLAN (Caring and Living as Neighbours)
- Credit Union Foundation Australia
- Daughters of Our Lady of the Sacred Heart Overseas Aid Fund
- Diaspora Action Australia
- Diplomacy Training Program
- Door of Hope Australia Inc.
- Edmund Rice Foundation (Australia)
- EDO NSW
- Engineers without Borders
- Every Home Global Concern
- Family Planning New South Wales
- Fairtrade Australia New Zealand
- Food Water Shelter
- Foresight (Overseas Aid and Prevention of Blindness)
- Fred Hollows Foundation, The

- Global Development Group
- Global Mission Partners
- Good Shepherd Services
- Good Return
- Grameen Foundation Australia
- Habitat for Humanity Australia
- Hagar Australia
- HealthServe Australia
- Heilala
- Hope Global
- Hunger Project Australia, The
- International Children's Care (Australia)
- International Christian Aid and Relief Enterprises
- International Needs Australia
- International Nepal Fellowship (Aust) Ltd
- International River Foundation
- International Women's Development Agency
- Interplast Australia & New Zealand
- Islamic Relief Australia
- KTF (Kokoda Track Foundation)
- Kyeema Foundation
- Lasallian Foundation
- Leprosy Mission Australia, The
- Live & Learn Environmental Education
- Love Mercy Foundation
- Mahboba's Promise Australia
- Marie Stopes International Australia
- Marist Mission Centre
- Mary MacKillop International
- Mary Ward International Australia
- Mercy Works Ltd.
- Mission World Aid Inc.
- MIT Group Foundation
- Motivation Australia
- Murdoch Children's Research Institute
- MAA (Muslim Aid Australia)
- Nusa Tenggara Association Inc.
- Oaktree Foundation
- Opportunity International Australia
- Our Rainbow House
- Oxfam Australia
- Palmera Projects
- Partner Housing Australasia
- Partners in Aid
- Partners Relief and Development Australia
- People with Disability Australia
- PLAN International Australia
- Quaker Service Australia
- RedR Australia
- Reledev Australia
- RESULTS International (Australia)
- Royal Australian and New Zealand College of Ophthalmologists
- Royal Australasian College of Surgeons
- Salesian Missions
- Salvation Army (NSW Property Trust)
- Save the Children Australia
- School for Life Foundation

- SeeBeyondBorders
- Sight For All
- So They Can
- Sport Matters
- Surf Aid International
- Tamils Rehabilitation Organisation Australia
- TEAR Australia
- Transform Aid International (incorporating Baptist World Aid)
- UNICEF Australia
- Union Aid Abroad-APHEDA
- UnitingWorld
- WaterAid Australia
- World Vision Australia
- WWF-Australia
- YWAM Medical Ships

Affiliate Members:

- Australian Federation of AIDS Organisations
- Australian National University – School of Archaeology and Anthropology, College of Arts and Social Sciences
- Charles Darwin University – Menzies School of Health Research
- Deakin University – Alfred Deakin Research Institute
- James Cook University – The Cairns Institute
- La Trobe University – Institute of Human Security and Social Change
- Murdoch University – School of Management and Governance
- Queensland University of Technology – School of Public Health and Social Work
- Refugee Council of Australia
- RMIT – Centre for Global Research
- Swinburne University of Technology Centre for Design Innovation
- Transparency International Australia
- University of Melbourne – School of Social and Political Sciences
- University of New South Wales-International
- University of Queensland – Institute for Social Science Research
- University of Sydney – Office of Global Engagement
- University of the Sunshine Coast – International Projects Group
- University of Technology, Sydney – Institute for Sustainable Futures
- University of Western Australia – School of Social Sciences
- Vision 2020
- Western Sydney University- School of Social Sciences and Psychology

* Denotes Interim Full Member

** Denotes Interim Affiliate Member