

Introduction – Towards Systemic Approaches to Evaluation and Impact

Barbara Befani, Ben Ramalingam and Elliot Stern

Abstract This *IDS Bulletin* is the second of two that follow an Institute of Development Studies event seeking to define an agenda for research and practice of development impact evaluation. It focuses on exploring the potential of systems ideas and complexity concepts to meet the increasingly complex challenges of an increasingly ambitious development agenda. In particular, the contributions seek to: (a) redefine 'learning' according to the number of 'learning loops' involved; (b) understand how to identify the most relevant impact evaluation questions; (c) simulate systems states in two sectors (leather and health) following the implementation of (combinations of) different policy options and other events; and finally, (d) shake the foundations of the impact evaluation institutional system, recommending that the notions of multiple perspectives and system boundaries are fully embraced, and that the system ultimately transitions from an 'evaluation industrial complex' to an 'evaluation adaptive complex'. While the issue is a step in the right direction, much more work remains to be done.

This special issue of the *IDS Bulletin* brings together two prominent themes in the evaluation of international development: what can be understood about the impact of development interventions, and the implications of development occurring within complex systems and settings. It therefore gives prominence to both systems thinking and complexity science, two perspectives that are increasingly drawn on by evaluators. This is the second of two *IDS Bulletins* following a workshop entitled 'Impact, Innovation and Learning: Towards a Research and Practice Agenda for the Future', held in March 2013 at the Institute of Development Studies.¹ This event brought together a distinguished group of international scholars and practitioners from academic institutions, donor country agencies, and multilateral development actors. It situated development evaluation in general, and impact evaluation in particular, in the specific setting of today's complex and changing international development context (see Befani, Barnett and Stern 2014).

In times of straightened budgets, accountability-driven demands to demonstrate the effectiveness of public expenditure are understandable. These are the circumstances that have led to the

emphasis on impact evaluations in international development. Policymakers are not only interested in 'evidence-based policy', but also wish to judge the effectiveness of specific interventions. How to achieve this level of clarity and measurement of policy impacts is especially challenging in today's development landscape.

This landscape has all the characteristics of a complex system. It is characterised by uncertainty as bilateral and multilateral aid donors are joined by new actors, including philanthropic foundations and other private donors, who play an increasingly important role. Resource flows are often unpredictable in an open and globalised world economy. The delivery of aid takes place within a broad setting of investment and trade, and is increasingly complicated by transnational issues such as migration, climate change and terrorism. Even providing a clear and complete definition of a development intervention is sometimes difficult (see Befani *et al.* 2014).

The 'map of development' keeps changing, with more types of interventions (country-led programmes and policies, public-private partnership projects, and civil society

development programmes, in addition to donor-funded projects) and a wider range of impact evaluation users (from donors and national governments to decentralised levels of government, non-governmental organisations, the private sector and communities) (Rogers and Peersman 2014). Additionally, the focus of development is now not merely on economic growth, but on growth that is equitable, socially inclusive, environmentally sound and that improves quality of life, taking many aspects of wellbeing into account (material, relational, perceptual). Many of these goals are hard to reach in the short term and some, like empowerment, require the combination of many factors and favourable conditions (see Befani *et al.* 2014).

In the search for evidence about ‘what works’ in this complex world, particular evaluation methods, like experiments and quasi-experiments, have been privileged (Gertler *et al.* 2011; HM Treasury 2011; USAID 2011) but at the same time have been hotly debated and critiqued from multiple perspectives: epistemological, methodological, practical and ethical (BetterEvaluation 2014; Bond 2014; UNEG 2013; Stern *et al.* 2012). A variety of weaknesses of experiments and quasi-experiments have been identified, which are detailed in Befani *et al.* (2014).

The event in March 2013 sought to enrich the debate by sketching out a research and practice agenda that would meet the increasing demands for new kinds of evidence about what works in development programmes and projects; evidence that would not sacrifice learning for accountability, that would speak to a broad audience including recipients and donors, that would capture the growing complexity of increasingly ambitious development goals, and that would fit within the multiple governance layers and lines of accountability of the new institutional settings of development assistance, post-Paris Declaration and post-Busan.

Issue 45.6 of the *IDS Bulletin* includes an overview of this agenda, equally split between methodological and institutional innovation. The current issue also addresses both methodological, or approach, innovation and institutional reform (or innovation), using the specific language and concepts of systems thinking and complexity science. These are the sources of many contemporary ideas around complexity,

interdependence, uncertainty, changing ‘boundaries’, emergent properties and the interaction of multiple actors and actions.

Systems thinking and complexity science draw on diverse roots – from the epistemological to the technological and mathematical. Although these perspectives are becoming well-established in some fields they are new to development evaluation, which has traditionally favoured more linear frameworks such as the much criticised ‘logframe’. Therefore, it is important to recognise that this issue of the *Bulletin*, more than the preceding issue, is exploring new frontiers in international development evaluation. The articles included here do not necessarily articulate a consensus or settled view among scholars or practitioners, and some of the ideas and methods advocated are very much ‘work in progress’. Furthermore, systems thinking and complexity science are very broad fields, and it has not been possible to achieve anything close to a full and comprehensive coverage of all the various traditions. Nonetheless, we believe that the articles in this issue make a useful contribution to an ongoing debate about how to assess effects and effectiveness without ignoring the complexity of the contemporary development landscape.

The ‘new methods/approaches’ agenda here includes four articles advocating changes at different levels: conceptual (stop thinking of ‘learning’ as such, and start thinking of ‘different types of learning’); approach (a framework to ask the right questions and refine them iteratively throughout the evaluation); and more strictly, methodological (modelling complex systems affecting – and that are affected by – interventions to refine our understanding of interrelationships and system dynamics in the theory of change). The two contributions advocating change in the institutional system suggest that institutions: (a) make boundaries of evaluation endeavours more explicit, or require evaluators to do so; (b) require that a multiplicity of perspectives are always taken into account; and (c) start taking steps towards becoming more of an ‘evaluation adaptive complex’ than an ‘evaluation industrial complex’ (what this means is further elaborated below).

Starting with the methods/approaches innovation articles,² Hummelbrunner uses systems concepts

to develop an 'exploratory' framework around learning in evaluation. He builds on existing learning typologies from educational psychology, neuropsychology and pedagogy, to argue that the general (and generic) concept of learning in evaluation (including impact evaluation) has little value, unless we make a distinction between at least three types of learning, distinguished by the number of 'loops' involved among different factors describing the learning process. Depending on the number of loops involved, learning can mean choosing the best option among alternatives (single-loop learning); changing the available alternatives and then selecting the best option (double-loop learning); and finally, designing a system to systematically set the boundaries of the available alternatives, selecting the available alternatives, and finally, picking the best option (triple-loop learning).

Hummelbrunner argues that different types of learning imply different evaluation questions: single-loop learning will be most common when the question is 'Did the intervention reach a particular goal?' (also simplified as 'Did we do things right?'); double-loop learning when the question is 'How can a particular goal best be achieved (or was achieved) under given circumstances?' ('Did we do – or are we doing – the right things?'); and triple-loop learning when the question is 'Who has (had) the power to decide how goals are (were) supposed to be reached?' (simplified as 'How do we establish rightness?'). The author then compares this typology of learning with types of value (instrumental, intrinsic, critical) and systems concepts (interrelationships, perspectives and boundaries).

The plea to broaden the questions asked by impact evaluations is reiterated by Garcia and Zazueta, who argue that asking the right questions requires the adoption of a systems perspective. This is because which questions are relevant only becomes apparent after the evaluation is already under way and fieldwork has begun, when often unexpected information becomes available. The authors, therefore, propose an iterative approach to designing impact evaluations where the specific objectives of the evaluation are likely to change in order to ensure that the ultimate objectives of the evaluation are met, in light of the continuous flow of relevant information. The authors argue

that understanding which questions are relevant is only possible if a thorough analysis of systems boundaries, components, interactions and emergent properties is conducted.

As with Hummelbrunner, Garcia and Zazueta's article is more about approach than about methodological choices: the authors provide examples of 'where to look' in order to frame the relevant questions. Analysing the boundaries of the system means identifying the relevant 'scale', which is often different than originally thought and should therefore be understood as a moving, rather than fixed, object. The components of the system are identified through the stakeholders that affect – and are affected by – the intervention. These are often more numerous and different than initial expectations. System interactions are understood to be the relations between the stakeholders, including interactions between the particular intervention being evaluated and past and present (concomitant) interventions. The authors show that an accurate analysis of these interactions can reveal unintended consequences; in the examples they provide, positive unintended consequences lead to an underestimation of the intervention's impact, while negative unintended consequences to its overestimation. In addition, focusing on the emerging properties of the system can help spot either transformational dynamics, when reinforcing causal loops are discovered (as in the ILUMEX case), or unwanted equilibria, with balancing causal loops (as in the South China Sea evaluation). The article can be considered a strong, evidence-based call for impact evaluation to shift from single-loop to double-loop learning.

The next two articles address the challenges of understanding how results are achieved in complex systems. This knowledge is often implicit in quasi-experiments, where control groups are constructed on the basis of hypotheses on 'confounding factors', which are causal factors able to influence the outcome of interest independently from the intervention. In modern development contexts, however, the claim that causal factors work independently as single causes, is becoming increasingly hard to justify. Derwisch and Löwe illustrate this with a system dynamics simulation of the leather shoes sector in Ethiopia, prepared for the United Nations Industrial Development Organization (UNIDO). Performance indicators like production levels,

jobs and wages are sensitive to specific interventions and contextual factors in complex and partially unexpected ways, to the point that the authors are motivated to identify not the single best policy option out of a number of alternatives (investment in equipment or skills, access to credit, strengthening the national quality system, the logistics, promoting labour standards, etc.) but the best mix, given the hypothesis of an external shock like the removal of trade barriers. This article should give readers an idea of how difficult reconstructing a counterfactual really is in a complex system, where multiple factors working as a 'package' influence the outcome in indirect and non-linear ways. Many alternative counterfactuals are equally plausible, and to select only one winner would require a near-impossible depth and breadth of prior knowledge. On the other hand, understanding the system's dynamics enhances learning about what systems states are possible or likely in the future, as well as helping explain current systems states on the basis of past events.

Along similar lines, Grove presents an application of a system dynamics simulation tool to the evaluation of the impact of a national strategy for the scale-up of HIV antiretroviral therapy. The simulation aimed to predict future systems states and allow decision-makers to learn about the need for different types of specialised staff at given timeframes, as well as the impact of given changes in the institutional structure of the HIV protection system in Zambia. It showed that unless some decisions were taken within specific timeframes, the burden of the disease could become uncontrollable and potentially catastrophic. The author regrets that, whilst the tool was well received and represented a shared and validated view of how the system actually worked, the implications of this knowledge for re-programming the strategy were not taken as seriously and the tool ultimately failed to realise its transformative potential. At the same time, the potential feasibility, acceptance and general added value of system dynamics modelling to impact evaluations is well illustrated, along with suggestions for how to make these more useful.

Within the 'institutional innovation' strand, Williams suggests that there are different levels of 'depth' with which institutions can embrace ideas from the systems and complexity traditions in order to reform the way impact evaluations are

done. He argues that if institutions merely require theories of change to be better developed in terms of causal interactions, allowing for non-linear relationships to be mapped and increasing uncertainty to be incorporated, the change will be welcome and even necessary, but remain somewhat prosaic. Williams argues that the full acknowledgement of perspectives and boundaries is key for these ideas to *profoundly* affect how impact evaluations are conducted. In practice, this means that institutions require a full analysis of the stakes of a wide range of stakeholders involved in interventions (the reasons why they care about the intervention and the evaluation; their motives, interests, preferences, goals, agendas, resources, ideologies). Embracing perspectives to the full means that institutions require evaluators to present different ways that the intervention can be framed and understood, by different people or groups; and in particular, what goals are relevant for whom, amongst those that the intervention can potentially contribute to. Williams then explains that becoming aware of boundaries means that institutions critically address issues of scale (which scales are relevant, which are not, and why); measurement (how targets and value standards are calibrated, and what the implications are for whom); decision-making (who is included and who is not); expertise (who is trusted, what knowledge is considered reliable, and what the consequences are for whom); and finally, legitimacy (what is considered 'right', according to whose worldview, are different worldviews included?).

In the concluding article, Reynolds applies the iron triangle metaphor, normally used to describe systemic failure (like in the post-First World War Paris Peace Treaty of 1919 or the resignation of United States President Eisenhower in 1961), to the current 'evaluation complex' within which impact evaluations are commissioned, conducted and potentially used. Reynolds suggests that the three vertices of the triangle are:

- (a) commissioners (decision-makers), who commission evaluations on the basis of their own predefined set of boundary judgements (on the object of evaluation, questions, methods, etc.);
- (b) administrators or implementers of interventions, who spend public money and put in place the activities eventually evaluated, trying to follow the cause-and-effect relationships outlined in the theories of change; and
- (c) the evaluators, who mediate different interests and perspectives.

Reynolds strongly criticises current interactions between these three ‘corners’ of the triangle, distinguishing in total six types of activities: auditing, planning, evaluating as understanding, evaluating as practice, commissioning, and learning from evaluations. According to Reynolds, auditing fails to appreciate situations as inherently or partly complex, and hangs on to the belief that some situations can be fully ‘tamed’ (through simple or complicated frameworks). He suggests that planning stops being ‘purposive’ (based on fixed goals and targets) and becomes ‘purposeful’ (based on agile measures of success, adaptable to changing situations); that evaluating as understanding abandons a positivist epistemology that seeks objectivity and bias avoidance, and adopts a constructivist epistemology with self-awareness of bias. Evaluation as sometimes practised may need to revise the ‘horses for courses’ belief that there is always a tool that fits best under specific circumstances, and craft new adaptations of existing tools and ideas in the face of emerging situations. Commissioning needs to broaden notions of rigour and conceive of ‘external coherence’ in terms of social and ecological responsibility. Finally, evaluations should serve not just as single-loop learning devices but become open to describing higher degrees of complexity and even conflict through double- and triple-loop learning. An evaluation system transitioning to these states could be called an ‘evaluation adaptive complex’, as distinct from the current system which, Reynolds suggests, could be more suitably labelled as an ‘evaluation industrial complex’.

In conclusion, according to several proponents of systems thinking and complexity science, impact assessment defined as net effect, quantitative measurement is simply not possible or desirable in complex social systems. However, with broader definitions of impact, and a plurality of approaches, methods and perspectives informed

by complexity and systems, there are considerable opportunities for rethinking the design and implementation of development intervention impact evaluations, well matched to today’s increasingly complex development landscape.

As noted previously, this issue only expected to offer a partial view of how complexity and systems thinking could inform impact evaluation. For a start, the articles here only cover a very small area within the broad school of complex systems research, and more work is needed to engage more fully with the conceptual and methodological possibilities that this area of work holds for evaluation.

More testing and applications of relevant frameworks and models in real-world settings are needed; the language, as well as the theoretical concepts, need to be more strongly connected to evaluation practice. There is scope for the currently growing body of knowledge and applications to be better documented and systematised.

At the same time there remain major practical questions and implications – for evaluation policies, systems, cultures and skills. Some of the applications presented here were only marginal components of much larger evaluations, or conceptual explorations. There is scope to bring complex systems ideas and concepts more fully into the various stages of impact assessment work, from design through to *ex post* evaluation.

If systems and complexity ideas are to be incorporated into the impact evaluator’s toolkit, much more work remains to be done. In the meantime, we hope that this issue of the *IDS Bulletin* represents a useful step in the right direction, and that its methods and agendas will stimulate insightful, conceptually sophisticated as well as practice-grounded debate.

Notes

- 1 The first of the two issues is: Befani, B.; Barnett, C. and Stern, E. (eds) (2014) ‘Rethinking Impact Evaluation for Development’, *IDS Bulletin* 45.6.
- 2 The first two articles in this category are not exclusively about methods, but rather about

new approaches or new framings of evaluative thinking. Through changing the way we formulate evaluative questions, these new conceptualisations should eventually influence methodological choice.

References

- Befani, B.; Barnett, C. and Stern, E. (2014) 'Introduction – Rethinking Impact Evaluation for Development', *IDS Bulletin* 45.6: 1–5
- BetterEvaluation (2014) *Rainbow Framework Overview*, www.betterevaluation.org/plan (accessed 4 September 2014)
- Bond (2014) *Evidence Principles*, www.bond.org.uk/effectiveness/principles (accessed 11 November 2014)
- Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B. and Vermeersch, C.M.J. (2011) *Impact Evaluation in Practice*, Washington DC: World Bank
- HM Treasury (2011) *The Magenta Book: Guidance for Evaluation*, London: HM Treasury, UK Government
- Rogers, P.J. and Peersman, G. (2014) 'Developing a Research Agenda for Impact Evaluation in Development', *IDS Bulletin* 45.6: 85–99
- Stern, Elliot; Stame, Nicoletta; Mayne, John; Forss, Kim; Davies, Rick and Befani, Barbara (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, London: Department for International Development
- UNEG (2013) *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, New York NY: United Nations Evaluation Group
- USAID (2011) *USAID Evaluation Policy: Evaluation, Learning from Experience*, Washington DC: United States Agency for International Development