
Barbara Befani

August 2013
The IDS programme on Strengthening Evidence-based Policy works across seven key themes. Each theme works with partner institutions to co-construct policy-relevant knowledge and engage in policy-influencing processes. This material has been developed under the Policy Anticipation, Response and Evaluation theme.

The development of this material has been led by IDS who holds the copyright.

The material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government’s official policies.

AG Level 2 Output ID: 507
Contents

Introduction 1

1 Background 3

2 Objectives 4

3 Participants 5

4 Programme 6
  4.1 Day One 6
  4.2 Day Two 7

5 Workshop Highlights 8
  5.1 Describing and explaining facts 8
    5.1.1 Key discussion points 8
    5.1.2 Key action points 9
  5.2 Engaging with (different) values and (contrasting) perspectives 9
    5.2.1 Key discussion points 9
    5.2.2 Key action points 10
  5.3 Improving management of impact evaluations and supporting learning use 10
    5.3.1 Key discussion points 10
    5.3.2 Key action points 11

6 Conclusions: building an agenda for the future 13
  6.1 Nine action points for a research agenda 13
  6.2 Nine action points for a practice agenda 14
Introduction

This report brings together notes and highlights from the International Workshop organised by the Institute of Development Studies (IDS) in Brighton on 26–27 March 2013. The event served as a launch platform for the Centre for Development Impact (CDI), a joint venture between IDS and Itad.

The main focus in this workshop report is to share background information about the workshop (participants, programme), as well as its purpose and highlights from the technical discussions. In particular, the latter signposts issues concerning both current practice and policy dilemmas, including areas where further thinking and innovation is needed. The report is written to stimulate thinking and questions for further work, and provides key pointers on an emerging agenda.

The note does not summarise papers presented at the Workshop, all of which have been circulated and are available online at:
http://www.ids.ac.uk/events/impact-innovation-and-learning-towards-a-research-and-practice-agenda-for-the-future

---

1 Although the note draws on contributions from all of those who attended the workshop, it has been prepared by the organising committee and does not claim to be a consensus view.
1 Background

The workshop arose from two sets of concerns: (1) those of evaluators who need a broader toolkit than the one currently available to assess development impact; and (2) those of commissioners who are worried about the inability of current evaluations to tackle impact questions appropriately and rigorously. Surprisingly, extensive developments in social science methodology over the last twenty years do not seem to have trickled down to the toolkits of development impact evaluators in a way that helps addressing these concerns; nor have new, more evaluation-specific methods made considerable contributions in this direction so far.

Common challenges that did not seem to be addressed adequately concern scientific standards like validity, replication and generalisation; the often unresolved conundrum of causal inference and attribution; contexts where long-term systemic transformations are influenced by a multiplicity of factors interconnected in complex, emergent and often unpredictable ways; and multiple value systems where contrasting perspectives offer different definitions of success. These challenges are partly scientific (i.e. related to the robustness and validity of evidence), and partly institutional and organisational (that is, related to ‘who learns’ and who gets to produce, synthesise and disseminate impact evaluation knowledge).

Against this background the Workshop was organised around four themes:

1. **Systems thinking**: how do we conceptualise systemic change and transformations and use systems in practice?
2. **Complexity**: which concepts and tools from complexity science do we need and how can they be used?
3. **Scientific quality standards**: which methods do we mix and why in order to maximise different types of validity and opportunities for replication?
4. **Causal attribution and contribution**: how can we use different models of causal inference to pick up clues, patterns, accounts and other evidence of impact?

The workshop was invitation-only: a limited number of key stakeholders were invited to actively participate in plenary sessions, paper sessions, panel sessions and group discussions. In order to maximise opportunities for discussion and network building, paper sessions were limited to 40 minutes and were followed by group discussions. At the end of group discussions, groups reported key discussion points to the plenary.
2 Objectives

The underlying intention was to locate the methodological debate within the broader transformation in aid architecture post-Paris declaration, whilst acknowledging the diversity of programmes and the plurality of impact questions that commissioners are interested in, and focusing on use: in particular on ‘who learns what, when’. This challenge was taken up particularly by the two keynote speakers, Professor Bob Picciotto (King’s College London, EES, UKES) and Professor Patricia Rogers (Royal Melbourne Institute of Technology). The presenters of all the papers were also asked to acknowledge the above cross-cut themes in their presentations, with the aim of concluding the event with an emerging consensus on answers and implications, highlighted in the final round table discussion with commissioners of impact evaluations from UNICEF, DANIDA, the World Bank and UNDP.

In selecting the participants, one goal was to reflect a broad range of stakeholder groups: academics, researchers, evaluators, independent consultants, clients and commissioners, donors, and users; reflecting the commissioning, conduct, theorisation and utilisation of impact evaluations in development. The workshop was conceived as an initial step in a longer-term process of engagement, and participants were invited to reflect on how they might shape the future agenda for research and practice in this area.
3 Participants

Paul Balogun (Balogun)
Chris Barnett (ITAD, IDS)
Julian Barr (ITAD)
Kelly Beaver (IPSOS)
Barbara Befani (IDS)
Kate Bingley (IDS)
Bernward Causemann (Causemann)
Gerry Bloom (IDS)
Laura Camfield (University of East Anglia)
Robert Chambers (IDS)
James Copestake (University of Bath)
Sabine Dinges (GIZ)
Marie Gaarder (World Bank)
Oscar Garcia (UNDP)
John Grove (Gates Foundation)
Richard Hummelbrunner (ÖAR)
Colin Jacobs (UKES, British Council)
Colin Kirk (UNICEF)
Andrew Lawson (Fiscus)
Michael Loevinsohn (IDS)
Peter Loewe (former UNIDO)
Richard Longhurst (IDS)
Sam MacPherson (ITAD)
Edoardo Masset (IDS)
Bruno Marchal (Institute of Tropical Medicine, Antwerp)
John Mayne (independent consultant)
Rosemary McGee (IDS)
Tamlyn Munslow (IDS)
Richard Palmer-Jones (University of East Anglia)
Bob Picciotto (King’s College)
Derek Poate (ITAD, UKES)
Ben Ramalingam (ODI)
Martin Reynolds (Open University)
Jennie Richmond (Oxfam)
Dane Rogers (ITAD)
Patricia Rogers (RMIT, Better Evaluation)
Cathy Shutt (University of Sussex)
Elliot Stern (University of Bristol)
David Todd (IDEDS)
Jos Vaessen (UNESCO)
Patrick Ward (Oxford Policy Management)
Bob Williams (Bob Williams, independent consultant)
Ole Winckler Andersen (DANIDA)
Aaron Zazueta (GEF)
4 Programme

4.1 Day One

8:45–9:45 Registration

9:45–10:15 Welcome and Introduction
Barbara Befani (IDS, CDI, EES), Chris Barnett (ITAD, IDS, CDI), Elliot Stern

10:15–11:15 **Keynote 1: Are development evaluators fighting the last war?**
Bob Picciotto (King’s College, UKES, EES)

11:15–11:45 Coffee break

11:45–12:25 **Paper Session 1: Why Systems-Based Impact Evaluation?**
(Chair: Elliot Stern, Editor of *Evaluation*)

Bruno Marchal (Institute of Tropical Medicine, Antwerp)
*Conceptual distinctions: Complexity and Systems. Making sense of Evaluation of Complex Programmes*

John Grove (Gates Foundation)
*Aiming for utility in 'systems-based evaluation': a research-based framework for practitioners*

12:25–12:45 Group discussion

12:45–13:30 Questions and comments from the groups

13:30–14:30 Lunch

14:30–15:10 **Paper Session 2: Applications of Complex Systems in Evaluation**
(Chair: Ben Ramalingam, ODI)

Aaron Zazueta (Global Environment Facility)
*Applying Complex Systems Theory in Impact Evaluations: a case study on the impact of GEF support in the South China Seas and adjacent areas*

Peter Loewe (UNIDO)
*Exploring the Potential of Systems Dynamics Modelling*

15:10–15:30 Group discussion

15:30–16:15 Questions and comments from the groups

16:15–16:45 Coffee break

16:45–18:15 **Panel Session 1: Values, Learning and Systems: a framework for critical rigour in impact evaluations**
Martin Reynolds (Open University), Bob Williams, Richard Hummelbrunner (ÖAR)

20.00–22.00 Buffet Dinner at ‘Alfresco’, Brighton beachfront
4.2 Day Two

9:00–10:00  **Keynote 2: Towards a research agenda for impact evaluation of development**
Patricia Rogers (Royal Melbourne Institute of Technology)

10:00–10:30  Coffee break

10:30–11:10  **Paper Session 3: Scientific Standards and Rigour**
(Chair: Barbara Befani)

  *Laura Camfield (University of East Anglia)*
  *Qualitative methods in Modern Impact Evaluation: Ensuring Rigour and Reliability*

  *Richard Palmer-Jones (University of East Anglia)*
  *Impact Evaluation, Replication and Ethics*

11:10–11:30  Group discussion

11:30–12:15  Questions and comments from the groups

12:15–13:15  Lunch

(Chair: Chris Barnett)

  *John Mayne*
  *Making Causal Claims*

  *Barbara Befani (IDS)*
  *Set-theoretic, diagnostic and Bayesian approaches to impact evidence*

13:55–14:15  Group discussion

14:15–15:00  Questions and comments from the groups

15:00–15:30  Coffee Break

15.30–16.00  **Bridge: Towards a more ‘impact-oriented’ institutional M&E system: common challenges and potential solutions from a UN perspective**
Jos Vaessen (UNESCO), Oscar Garcia (UNDP)

16:00–17:30  **Panel Session 2: How to create the conditions for innovative research and practice in impact evaluation**
Elliot Stern (Chair), Colin Kirk (UNICEF), Ole Winckler Andersen (DANIDA), Marie Gaarder (World Bank), Oscar Garcia (UNDP)

17:30–18:00  Concluding remarks from all: what have we learned?
Elliot Stern (University of Bristol), Barbara Befani (IDS) and Chris Barnett (CDI)
5 Workshop Highlights

What follows synthesises the main discussion and action points emerged from the event, summarised around three main themes: (1) describing and explaining facts; (2) engaging with different values and contrasting perspectives; and (3) improving management and supporting learning and use.

5.1 Describing and explaining facts

Impact Evaluations must check the quality and scientific rigour of claims of factual evidence, and develop new ways of describing and explaining facts, by developing better Theories of Change, but also accepting that facts aren’t always fully predictable or knowable. There is a need to engage with Systems Thinking in all its forms (Complex Systems, Complex Adaptive Systems and Critical Systems Thinking). In terms of outcomes and impacts, a more decisive and robust shift was advocated from short to long term, from positive to negative, from anticipated to unexpected, and from ultimate to intermediate results.

5.1.1 Key discussion points

1. There is a theory–practice gap in the quality of evidence. Not all research quality criteria are ensured or even discussed in the presentation of research findings: all methods require assumptions but most often these assumptions are not tested.
2. We need better theories of change and theory-based evaluation, in particular realist evaluation: under what conditions does a programme work and why?
3. Most programmes are either complicated, complex, or both; and are usually not evaluable by experimental methods: ‘we have a very good hammer for an increasingly small nail’. Society is not a laboratory: ‘poverty is not a disease and development aid is not a drug’.
4. We need to understand interrelationships between programme components as well as between the programme and the wider system. A systemic effect arises from the interaction of parts, and the whole is more than the sum of its parts.
5. The traditional logic model was heavily criticised for its linearity. Causal chains are too simplistic and unable to account for the variety of possible outcomes. Logic models can work if a number of assumptions hold, but these assumptions are usually not transparent nor built into the model.
6. Complex systems are useful because they are made of multiple interconnected elements, nonlinear interactions, non-proportional effects, negative and positive feedback loops, time delays, path dependence.
7. In some cases change can only be understood holistically: relations and history matter.
8. We don’t always know what we don’t know; in addition to ‘known unknowns’, we also work with ‘unknown unknowns’. In these cases, hindsight does not lead to foresight; summative evaluation may be necessary but not sufficient to predict the future; and what worked in the past may not work in the future.
9. Development interventions don’t work on their own, but neither ‘in addition’ to other causal factors; most often they combine with other factors and influence the outcomes in ways that do not add up to each to other, but are relatively unpredictable and context specific. Interventions might be triggers, or might prepare the ground for other causes to trigger the outcome in the future; or yet again, might act as sustaining support for continued effect.

---

2 From here onwards, unattributed comments in speech marks are from workshop participants.
5.1.2 Key action points

1. Use ethical codes, peer review, data archiving, restudies, and more publications of negative or null results, in order to improve the quality of evidence.
2. Use theories that identify stakes, not just stakeholders; and paint a picture of the stakeholders’ incentive structures (motivations, drivers). Theories of change need to include external supporting factors, which are what is needed in addition to the intervention to achieve impact. Theories of change must also be evolving and adaptive – change as you go along, particularly in emergent settings.
3. Use different approaches and skills depending on whether you are dealing with complicated or complex aspects of interventions. For example, in complex settings we need adaptability, flexibility, inclusion of multiple perspectives not just the evaluator’s; piloting and testing, inter-multidisciplinary teams and participatory approaches.
4. Use feedback ‘loops’ instead of linear ‘chains’.
5. While counterfactual analysis and regularity-based models have their place within the evaluator’s toolkit, more use should be made of other approaches to causal attribution, like configurational and mechanism-based frameworks.
6. Design a process whereby possible explanations/causes are confirmed or rejected explicitly. Look for both evidence that supports/confirms your assumptions and evidence that weakens/disconfirms rival explanations. Think in terms of necessity and sufficiency.
7. Accept uncertainty and, sometimes, unknowability. If patterns are emergent and unpredictable, start from scratch. In these cases favour participatory over expertise-based approaches. When there is no obvious roadmap, do bricolage. Use flexible and adaptive designs that allow learning, and a wide range of observation and collection methods, like process analysis, historical methods, and real time evaluation.

5.2 Engaging with (different) values and (contrasting) perspectives

Impact evaluations are not just about describing, explaining and attributing outcomes to interventions or other causal factors; like all evaluations they should be centred around the attribution of merit worth and value. Impact is not just about effectiveness but also relevance. How can implicit values about results, processes and distribution of benefits be made explicit? Different stakeholders might have different values and perspectives. How do we reconcile them? How do we synthesise different values into an overall value judgement?

5.2.1 Key discussion points

10. The role of a wide range of stakeholders is reinforced by emergence and (unknown) unknowns in processes and outcomes; by the principles of the Paris declaration, and by contemporary theories of policymaking and public management that blur the distinction between policymakers and citizens as well as between policymaking and policy implementation.
11. Systems are not about the best possible representation of reality, but about the conversation amongst stakeholders on what the best representation is. The value of the model is in the response and engagement it generates. The point is not the model, but the modelling process: modelling is, ultimately, a meta-language.
12. Different types of learning were discussed. Single loop learning (popular in audit and control systems) deals with interrelationships and instrumental values: what are the steps to reach a goal? The goal is not questioned. If an intervention doesn’t work, it gets changed: learning leads to adaptation. There is one best way of doing things, and we need to learn the right way, asking ourselves ‘are we doing things right?’
13. In double-loop learning, you have an overview of the mechanisms that make things work in given contexts. You don’t try to solve problems as they arise, but investigate the underlying causes and assumptions, and how the consequences fit with the different values of the actors involved. Purpose and goals get questioned. Different framings for the same intervention are developed and informed by different intrinsic/personal or organisational values. The question is, ‘are we doing the right things?’

5.2.2 Key action points

8. Incorporate multiple (even contrasting) perspectives, and get to know stakes and the incentive structures of stakeholders (motivations, drivers, along with values).
9. Tweak well-known methods to focus on the value dimension: we don’t necessarily need new methods to incorporate values into impact evaluation.
10. Challenge the idea of a single system when using systems ideas to aid the valuing process: recognise that all systems are partial.
11. Use double-loop learning.

5.3 Improving management of impact evaluations and supporting learning use

Impact Evaluation is not just about describing or explaining facts and assessing the value of interventions. The latter is done for a reason: learning, accountability or more generally ‘use’; and the process is managed by institutions, authorities and agencies within a global community of actors. Can agencies create the conditions for better impact evaluation?

5.3.1 Key discussion points

14. The commissioning and management of impact evaluations takes place in a political space built on opportunities and constraints. The boundaries of this space should be explored explicitly: who decides which framework is more important? What is in and what is out? Who is affected and who isn’t by the choice of perspectives? What makes these decisions the right thing to do? Who says that a given choice is the right one?
15. What is the best process to develop an evaluation design? Does the design need to be part of the Terms of Reference (ToR), does it need to be developed by the external evaluators in their proposal, or does it need to be part of a separate project? There is no evidence about what the best process is to develop an impact evaluation design.
16. It is much easier to evaluate interventions that directly target beneficiaries, in contrast to focusing on institutional capacity building which is a major part of the UN portfolio. Most interventions are multi-actor, multi-stranded and multi-sided. What is the scope for deconstruction? Does it make sense to isolate and evaluate single interventions components or do programmes need to be evaluated holistically?
17. Is the evaluator an ‘activist inspired by ethical imperatives’ or a professional responding to an incentive structure? In this context the issue of the ‘enabling environment’ has been raised: managers need to be trained to understand and use results presented in a less simplistic, more sophisticated and nuanced (and perhaps equity-focused) way, going beyond average effects, logic models and linear chains that characterise Results Based Management.
18. In a context of decentralised power, governments listen more carefully to communities and so should evaluators: partnerships have a fundamental role.
19. Evaluation is a peculiar market with several information problems, where commissioners don’t know in advance what they are buying and sometimes don’t even know what they want to buy.
20. If aid is meant to be harmonised, so should processes connected with evaluations as far as possible: for example, data collection and results presentation. One issue to consider is sharing information and avoiding redundancy while respecting privacy.

21. Innovation is not a goal in itself: it must be a means towards an end. Learning and use is the ultimate goal of evaluation: what innovative practice can we foster in order to improve ‘the impact of impact evaluations’? Quality does not necessarily imply use.

22. Is it possible to have accountability without learning and learning without accountability? The more radical notion of accountability assumes that, ‘if you haven’t shown impact it means there hasn’t been any; there is no benefit of the doubt’. Experimental methods have been said to be ideal for accommodating this radical notion of accountability, and this has been mentioned as the probable reason behind the gigantic divide between the amount of funding channelled into experimental methods and the funds allocated for other methods.

23. In less radical notions of accountability, policymakers must not only learn to be accountable for achieving results and having an impact, but also to be accountable for learning. In some circumstances learning could be a result in itself, an intermediate outcome, and ultimate impact could be impossible to achieve without a good amount of learning first. But how do we measure learning? Can learning be proved through action? What types of learning shall we consider? Whose learning is taken seriously?

5.3.2 Key action points

12. Conduct more research on independence. What happens when evaluators and commissioners have contrasting values? What is ultimately needed in order to protect ‘those speaking truth to power’? Is it specific personality characteristics like independence of mind, or specific institutional and organisational structures? Or perhaps skills such as facilitation and critical thinking?

13. Improve evaluation quality through public validation of what constitutes better or weaker aid evaluation; reputation mechanisms; investment in longer and more flexible contracts with evaluators that include building, writing and disseminating innovative methodologies; in-depth studies of the factors and barriers that foster or inhibit innovative practice (e.g. multidisciplinary and transdisciplinary practice); and understanding the incentive structures of the actors operating in the evaluation business.

14. Support communities to have genuine involvement in decision-making, beyond mere involvement in data collection.

15. Conduct more research on the political economy of the evaluation market, in order to explain why quality is often low.

16. Set up a strategic discussion on strengthening evaluation systems (including for example what data to collect, how to structure it, record it, report it and ultimately make it comparable), that would facilitate the harmonisation of data collection systems and allow organisations to share data while respecting privacy and avoiding redundancy.

17. Manage expectations on use. For example, ‘we should not expect direct instrumental use but rather indirect and cumulative learning that stems from gradual changes in the evaluation culture’.

18. Promote a blurred distinction between accountability and learning, and avoid falling into the trap of, ‘learning to do the wrong things righter’.

19. Be more humble and recognise that all science is prone to bias; more energy should be devoted to making bias transparent rather than minimising bias at all costs.

20. Conduct more research on the relation between evaluation quality, uptake and development.
21. Engage in ‘triple-loop learning’: that is, learning not just to change, but learning to learn; not just reflecting on the rules that govern and influence behaviour, but also trying to change those rules. The reflection on boundaries should explore who determines the goals and purposes, what we consider as possible choices: do choices stem from power, from path dependence, or from authority and knowledge? Are choices legitimate? Triple-loop learning is about the political nature of ethics and about the critical/political values that explain why a certain frame has been chosen or imposed. Whose values finally count? Critical Systems Thinking has been offered as one suitable approach for triple-loop learning as it also deals with ethics and power.
6 Conclusions: building an agenda for the future

The following points are meant to trace a roadmap for a research and practice agenda as it emerged from the event. They synthesise the action points outlined above, and are grouped together under the two ‘research’ and ‘practice’ headings, although most points crosscut the two domains.

6.1 Nine action points for a research agenda

a) **Devote more energy to making bias transparent** rather than expecting to avoid bias at all times; recognise that all science is prone to bias.

b) **Conduct more research on independence.** What happens when evaluators and commissioners have contrasting values? What is ultimately needed for independence, in order to protect ‘those speaking truth to power’? Do we need specific personality characteristics, organisational structures or skills?

c) **Conduct more research on the political economy of the evaluation market,** and try to explain why quality is often low.

d) **Conduct more research on the relationship** between evaluation quality, uptake and development.

e) **Improve the quality of impact evidence,** through: ethical codes; peer review; data archiving; restudies; more publications of negative or null results; public validation of what constitutes better or weaker aid evaluation; reputation mechanisms; investment in longer and more flexible contracts that include building, writing and disseminating innovative methodologies; in-depth studies of the factors and barriers that foster or inhibit innovative practice; and more generally on the incentive structures of the actors operating in the evaluation field.

f) **Improve harmonisation of data collection systems:** start a strategic discussion on what data to collect, how to structure it, record it and report it, while respecting privacy and avoiding redundancy.

g) **More non-experimental causal inference.** While counterfactual analysis and regularity-based models have their place within the evaluator’s toolkit, more use should be made of other approaches to causal attribution, like configurational and mechanism-based frameworks. Design a process whereby possible explanations/causes are confirmed or rejected explicitly: look for both evidence that supports/confirms your assumptions as well as evidence that weakens/disconfirms rival explanations. Think in terms of necessity and sufficiency.

h) **Use better theories of change:** identify stakes, not just stakeholders; provide an overview of the stakeholders’ incentive structures (motivations, drivers); include external supporting factors that are not affected by the programme.

i) **Use flexible and adaptive designs** that allow methods to change in response to learning and a wide range of observation and collection methods, such as process analysis, historical methods and real-time evaluation.
6.2 Nine action points for a practice agenda

a) **Manage expectations on use**: expect indirect and cumulative learning that stems from gradual changes in the evaluation culture rather than direct, instrumental use.

b) **Blur the distinction between accountability and learning**, and avoid falling into the trap of ‘learning to do the wrong things righter’.

c) **Support communities** to have genuine involvement in decision-making.

d) **Incorporate multiple (even contrasting) perspectives**, getting to know stakes and the incentive structures of stakeholders (motivations, drivers, along with values).

e) **Incorporate values** into well-established impact evaluation methods.

f) **When patterns are emergent and unpredictable, start from scratch**, and at this stage favour participatory rather than expert-based approaches.

g) **Engage in ‘double-loop learning’**: question purpose and goals, and aim for an overview of the mechanisms that work in given contexts. Investigate underlying causes and assumptions and ask yourself, ‘are we doing the right things?’ in addition to, ‘are we doing things right?’

h) **Engage in ‘triple-loop learning’**: use Critical Systems Thinking to reflect on, and perhaps change, the rules and boundaries that influence behaviour and determine goals and purposes. Do choices stem from power, from path dependence, or from authority and knowledge? Are choices legitimate? Whose values finally count?

i) **Accept uncertainty** and sometimes unknowability.

The organising committee:
Elliot Stern, AcSS, University of Bristol
Barbara Befani, PhD, IDS
Chris Barnett, ITAD