Special Debate Section

Does Impact Evaluation in Development Matter? Well, It Depends What It's For!

Irene Guijt[a] and Chris Roche[b,*]

[a]Research Associate, Overseas Development Institute, Canberra.
[b]Institute for Human Security, La Trobe University, Melbourne.

*E-mail: c.roche@latrobe.edu.au

European Journal of Development Research (2014) **26,** 46–54. doi:10.1057/ejdr.2013.40;
published online 14 November 2013

Keywords: impact evaluation; accountability; impact monitoring; empowerment; rigour; complexity

Introduction

It's never fun to waste one's time. So an Impact Evaluation (IE) that makes a difference would appear to be a good starting point, right? Yet the utility of IE is not always clear or questioned. In a recent evaluation of 3ie (Morton *et al,* 2012), a strong advocate of quasi-experimental impact studies, the external evaluators noted the disappointing interest of the core target group – policymakers – in the studies that IE has to offer.[1] A recent conference on evidence-informed policymaking (Newman *et al,* 2013) concluded that the assumed utility of IE for policymaking needs serious questioning as little evidence exists on policymakers' capacity to use research evidence, with evidence instead of the insidious effect of corruption on use of evidence in policymaking processes. By contrast, at the recent Big Push Forward event on 'the Politics of Evidence', one panellist named eight randomised studies that had all been very useful. In this case, utility was defined for very different purposes and ranged from building support to expand existing programmes to checking policy relevance, understanding policy adaptations needed in other contexts and improving programmes. So why is IE useful in some cases but not yet in all?

Our central argument is the importance of refocusing the debate on the core functions and purposes of IE. This focus has implications for the 'who', 'what' and 'how' of IE (see Box 1).[2] It also has the advantage of steering us clear of the tedious method-wars that largely ignore the purpose question. In order to assess why, and indeed whether, IE is important, we need to be clear what it is meant to achieve. Our view is that this is too often forgotten in the heat of the IE methodological battles and 'paradigm wars'. Understanding the purposes of IE allows us to be

Box 1: Two types of impact questions (Pritchett *et al*, 2013)

- Funders' questions: allocating their resources efficiently requires measures of *what* different options can achieve.
- Implementers' questions: must learn iteratively which actions achieve the best results, in complex and often unpredictable ways, so they must learn *how* they do this.

clearer about the degree to which current approaches can contribute to each purpose, as well as to clarify what we might need to do differently in order for those purposes to be achieved.

So What is the Problem?

Many of the debates and tussles about IE focus on who is right or wrong in relation to which method, and on formal definitions of what 'impact' is, or is not. These tussles are embedded in a growing IE industry with large volumes of money and many professional reputations at stake, which also shape those exchanges. Tussles are sometimes fought in terms of method hierarchy, such as cost-benefit being better than Randomised Control Trials (RCTs), RCTs being better than asking citizens' experiences and so forth. Tussles are also fought by sighing over the supposed impenetrability of certain 'disciplinary codes' (Harrison, this volume).[3] Tussles can become meaningful discussions when there is more willingness and ability to bridge codes and remain curious. Judging people's arguments in terms of their ability to be translated for economists (sic) is certainly one option to 'bridge codes', as proposed by Harrison (this volume). However, it may just be a possibility that economics is not the only perspective that matters, a position we take.

Tussles such as these detract attention from the more important issue of the purposes of IE, and the implications for practice. Our experiences suggest that IE can contribute meaningfully to three purposes.

Learning to 'Improve' As Well As 'Prove' What Works, or What Does Not

Many organisations struggle on a daily basis to balance investing appropriately in ways that help show what strategy is 'working' (shorthand=prove), as well as in ways to ascertain how to ameliorate implementation (shorthand=improve). Although both are necessary, and indeed are linked, we believe there can be a tension between the two. The learning purpose requires us to get serious about not only *whose* views and perspectives need to be central to understanding and valuing impacts in different contexts, but also *whose* learning counts. Furthermore, the contingent nature of impacts also suggests that we need to be careful about the extent to which we can treat the past, or different environments, as a good indicator of 'the truth', particularly in highly context-sensitive interventions that are unlikely to be replicable. Furthermore, Woolcock (2009) cautions against inferring too much from data based on one comparative moment, given the different trajectories at which impacts emerge – hence the importance of aligning the timing of when insights are needed, when impacts emerge for those intended to benefit and when IE studies are undertaken.

Accountability for Funds Used

But how do we understand accountability? Following Brown (2007), it is simply the 'responsibility to answer for particular performance expectations to specific stakeholders'. What performance expectations and which stakeholders matter, and who answers to whom (see Box 2, Brown, 2008) are key questions.

Too often, accountability is dominated by the notion of contractual (principal–client) relationship, that is, how a 'donor' gets an organisation to which it gives funds to do what the donor wants. This is particularly problematic when precise objectives, or the means to achieve those objectives, are difficult to specify in advance (Shutt, 2012). Furthermore, in many IEs,

Box 2: Forms of accountability (adapted from Brown, 2008)

*Principal–Client Accountability – Client acts for principal through contractual incentives*Requires clarity of outcomes and means to achieve them established in advance.

- Benefits from objective measures of achievement or performance.
- Tends towards contractual agreements: sanctions are legal or economic.

Peer accountability – Performance is assessed by colleagues or experts based on agreed standards.

- Used when outcomes, or means to achieve them, are less easily specified as objective measures.
- Uses expertise, or intimate knowledge, to make judgements about quality, often based on inter-subjective agreement.
- Based on expert judgement: sanctions are reputational.

Social or mutual accountability – Partners act on shared goals backed by relationship and social identities.

- Requires negotiation between parties to establish 'rules of engagement' and 'what success look like' that is appropriate to context.
- Needs to create opportunities for different parties to learn and adapt.
- Tends towards compacts: sanctions are moral or social.

Political accountability – Representatives act based on mandate of constituents.

- Allows 'constituents' to periodically choose and remove those who represent or provide services to them. Requires transparency and freedom of information to function well.
- Constituents form own judgements based on own preferences.
- Based on mandates: sanctions are political, bureaucratic.

citizens intended to benefit – and in particular *their* measures of performance – are rarely included. This is the case even though their 'net welfare' can be included in the outcomes under scrutiny – and usually is.[4]

What form of accountability one values and seeks to uphold, what purpose this serves and for whom, needs to be discussed and deliberated within any IE process (see also Picciotto, this volume). The dominance of a 'principal–client' focus on accountability can block constructive thinking, for example about governance of international development (Booth, 2012). It also drives narrow understandings of how IE might be conceived and its utility. Principals are usually those commissioning an IE, and therefore determine what focus, methods and approaches they deem to be appropriate. Social or mutual accountability can generate space for negotiating towards an IE of mutual interest. Political accountability can help us imagine how an IE process might include a more robust input from citizens based on their preferences and interests.

Influencing for Empowerment

In addition to influencing the practice of agencies, IE studies have the potential for much wider use by holding others to account. The findings should at least be accessible and understandable to those who might want to scrutinise and contest the findings, or use them to promote better policy and practice. Moves towards greater transparency, for example through the International Aid Transparency Initiative, are particularly important in this regard, as are the growing number of collections of evaluation studies. Therefore, the process of IE and its findings need to be made available in order for data to be contestable and outcomes to be accessible – *and* usable – by communities and agencies to advocate for improved performance. We need to invest much more in ensuring that findings can be used by citizens whose welfare needs drive international

development, at least in theory. We will need to do much better than ideas such as those offered by Harrison (2014) to peruse published papers in – the often gated – major journals of economics. These will be of little help to the people he so stridently defends, given that most articles are in English and written in an inaccessible disciplinary code, even assuming that the prohibitive cost could be overcome.

Ultimately, as outlined above, IE should improve the effectiveness of international development through three routes: practice, policies and power.

Encouraging Better Practice

But how do we think this should happen, given the persistent difficulty to ensure useful information on the one hand and capacity for uptake on the other? Practice needs changing now – not later – and thus IEs need to generate *timely insights* in order to adjust the work *en route*. Yet many IE studies typically appear beyond the moment of corrective action. This means expanding the notion of IE to include impact-oriented monitoring (White, 2014) currently ruled out by orthodox approaches and definitions of both monitoring and IE.

Encouraging Better Policies

But how do we understand this given IE, thus far, has often not captured the attention of many policymakers, notwithstanding some notable and much-cited exceptions such as the study of the Progresa conditional cash transfer scheme (itself not without its critics)? Much of the literature on policy uptake, including the link between research and policy outcomes, provides useful lessons on increasing the likelihood of uptake. Design and implementation would do well to consider, for example: (1) recognising and working with the political and contested nature of any policy process as 'evidence informed' rather than 'evidence based'; (2) engaging stakeholders in the IE process as part of the exercise rather than simply submitting a report to them at the end; and (3) providing findings, insights and outputs in forms that are engaging and accessible for diverse audiences (notably for the 'influencing purpose' described above).

Changing Power Relations

But how can this be built into IE? If we recognise that IE processes are embedded in and subject to the broader political dynamics of any evaluation process (try applying for a grant without an RCT specialist on board, as Harrison writes, this volume), then understanding power relations within the field of IE is critical in order to shift these (Big Push Forward, 2013). The asymmetries in access to information and power between stakeholders in development are considerable. Think of the differences between a large data-rich Development Bank and small farmers receiving micro-finance from them. Ignoring these differences means that IE can simply uphold existing power inequalities. To be clear, we are not advocates of advocacy-driven research that seeks to discover findings simply to serve political ends. Instead, we would like to see more IE studies that are conscious of existing power differentials and seek to address them as part of the methodology (see Box 3). Practical steps for this include ensuring the IE process does not privilege certain voices and exclude others (or is clear about which voices it privileges and why), and making findings available *and accessible* in ways that do not further exacerbate asymmetries in information and knowledge.

Box 3: IFAD's participatory impact assessment and learning approach

In 2013, the International Fund for Agricultural Development (IFAD) with the Bill and Melinda Gates Foundation (BMGF) initiated the piloting of a new methodology for understanding impact in which the first author is closely involved. In the first pilot in Vietnam, the sampling frame focused on seeking the experiences of those in the official poorer and near-poor categories. Final village-level sessions were facilitated that led to intense debates between different socio-economic groups about who felt which impact and to what extent. These discussions, held in all 18 villages, have refined the initial village-level impact assessments, all of which will feed into a large participatory sensemaking session that includes provincial and district-level officials and project staff to conclude on key changes, underlying causal explanations and the project's impact, that is, contribution to improvements for local citizens. Documentation is in progress.

So What Matters?

A constructive debate about IE transcends disciplinary squabbles and methodological tussles about which method is best. What do we need to clarify when designing, implementing and using an IE? What matters in these debates? A thorough discussion about the assumptions we all bring to IE will ultimately help impact studies contribute to development effectiveness – bringing us to the importance of epistemology, how we know what we know (Roche, 2013). We offer five elements as a starting point for clarifying assumptions and broadening the scope of IE: standards, rigour/utility, power and politics, evidence and uncertainty.

Standards Matter

No one contests the importance of having standards by which an intervention or change process can be judged as being 'effective', or standards that can shape the methodology. However, different views are held in relation to the relative merits of 'imprecise accuracy' versus 'precise inaccuracy' – and therefore which standards are wielded when pronouncing on the validity of an IE. Many mainstream IE studies, including the noteworthy Progresa study (IFPRI, 2013), have been critiqued for their precise inaccuracy. More honesty is needed about what inference is possible for each study and across studies. Therefore, standards matter across IE design, implementation and use. Such a holistic perspective on standards and the trade-off between design, on the one hand, and, on the other, implementation and use is often ignored. Yet examples exist to help our thinking (see Box 4).

Rigour and Utility Matter

Rigour is one of the most contested aspects of quality within IE. In particular, certain views on statistical rigour in order to – mathematically at least – ascertain causality appear to have trumped other considerations (see Picciotto, this volume). The recent conference on the Politics of Evidence (see footnote 15) raised the need for 'relevant rigour' *and* 'rigorous relevance' when valuing evidence. This entails having a standard for rigour that is relevant to the purpose of IE, as well as being rigorous about ensuring that the IE is relevant. Linking relevance and rigour challenges the presumed ownership of the term 'rigour' by those advocating for (quasi) experimental methods. Professional scholarship is important, but this must extend beyond narrow versions of economics, or one-upping one method with another. Rigour needs to be negotiated – there is no single standard, one method that is inherently more rigorous than others, nor is it just a matter of rigorous measurement. Of course we need to be rigorous about how we ascertain causality, but we also need to be rigorous about assessing which domains of a given intervention

Box 4: The Rainbow Framework for evaluation standards (http://betterevaluation.org/plan)

One example of a holistic approach to evaluation, including IE, is the Rainbow Framework of the Better Evaluation collaboration. This framework argues that seven clusters of tasks need careful consideration to ensure the quality of any evaluation process:

- Management of valuation of evaluation system.
- Defining what is to be evaluated.
- Framing the boundaries for the evaluation.
- Describing what has happened and the context within which it occurred.
- Understanding the causes of what has happened.
- Bringing data and analysis together in an overall conclusion or judgement.
- Reporting and supporting use of findings.

Box 5: Multiple approaches for causal inference (Davidson, 2013)

- Asking observers or participants.
- Looking for 'signature' traces or 'distinctive effects' of a programme.
- Checking the relationship between 'dose' and 'response'.
- Making comparisons with another group.
- Controlling statistically for other variables.
- Identifying underlying mechanisms.

require knowledge and therefore merit attention. This kind of rigour makes possible different approaches to determining causal inference that are appropriate for the questions we are asking (see Box 5, Davidson, 2013). Which approach, or combination of approaches, adopted will depend on the questions being asked, the level of precision or certainty required, the available resources and stakeholders' interests? In short, rigour and relevance are intricately linked and determine the degree to which a study is fit for purpose and methodologically appropriate to the nature of the intervention and the context.

Power and Politics Matter

Power relations affect evaluative processes in many different ways: in the choice of what is evaluated, methodological choice, who evaluates, what is valued by whom, its eventual use – and above all who decides all of the above. The politics of evaluation are usually more intense than is the case for normal research, particularly when the findings can affect people's status, position and livelihoods. For impact studies to be useful, we need to better understand how evidence (see below) is used in policy processes, which are inevitably mixes of facts, values and political considerations (Prewitt *et al,* 2012).

Factoring power and politics into IE is important regardless of the method that is selected – we ignore it at our peril. One poignant example is the controversy associated with the Gates-funded HIV vaccine pre-exposure prophylaxis trials in Cameroon and Cambodia (McGrory *et al*, 2009), despite the rigour of the research method and its ethical clearance process. In this example,

unaddressed conflicts with HIV activists who had various concerns about the evaluative research led to the disruption of two expensive medical trials.

Evidence Matters

Evidence clearly matters. No one is against evidence. Caricaturing those who question the hegemony of RCTs as the best approach to IE as constituting an 'anti-evidence' movement (as one tweet suggested of some of the Big Push Forward convenors) is unhelpful. More helpful is to untangle what constitutes credible evidence for those who are supposed to use it, which in itself is not a new issue or debate (Claremont, 2006). Picciotto's article (this volume) highlights the problems with the experimental school of IE, which has a 'parsimonious' perspective on evidence, ignoring the results chain and focusing on simple attribution calculations. These studies, often one-off, offer some evidence – often average effects – to answer the question for policymakers of 'what works'; however, this narrow focus is critiqued by many contributors to this volume. The utility of the evidence they offer is also critiqued vigorously for other reasons (Deaton, 2013), not least for the focus on internal validity that hinders generalizable conclusions beyond the area of study (see Lensink this volume) – hence the current rush for replication studies and systematic reviews. On the other hand, we also have 'results' data. The current data greed emerging in many development bureaucracies that centres on the results of programmes and projects do not always lead to useful evidence of the pathways of change. When undertaken well, this kind of evidence can show what outputs have led to which outcomes, but rarely does this consider other explanations for the observed outcomes – and rarely is results data used to understand how change happens and why.

We see the need for different types and sources of evidence – depending on the question, the type of intervention and the purpose. Two other forms of data can be used to complement the big E (impact) and the small E (results) data. First, evidence that tracks the *emerging* impact of programmes as they are implemented, and adapted. This evidence is far less accepted, yet, as argued above, could make a crucial contribution to make IE more useful for effective development by providing ongoing feedback. Second is the recent emphasis on 'big data', that is, from more direct forms of citizen feedback encouraged by organisations such as Ushahidi and Twaweza, and real-time data collection from multiple sources through, for example, the UN Global Pulse. These examples point to the utility of diverse sources of information by highlighting what they enable:

- Direct feedback from those whose programmes seek to benefit so they can publicly comment on the agencies implementing or funding these programmes, hence shifting power relations (to some degree) between actors.
- The possibility for cross-checking and triangulating impact data from multiple sources and through different media.
- Facilitating the identification of unexpected correlations, for further investigation and testing of causal links.

Uncertainty and Complexity Matter

Although not everyone has signed up to the notion that development work is largely focused on addressing 'wicked problems',[5] there is emerging consensus that development involves systemic change, which, by definition, is not linear, nor predictable (cf Beinhocker, 2007, Ramalingam and Jones, 2008, Barder, 2012). Hence, understanding patterns and correlations that can reduce

uncertainty is as important in many contexts as proving causality. Furthermore, there is a renewed understanding of the difference between 'risk' and 'uncertainty' in many areas. As Andrew Haldane of the Bank of England has suggested (Davies and Haldane, 2012) in relation to the global financial crisis, part of the problem was *driven by the quest for certainty, and mathematisation of economics was a means of achieving that certainty*. An alternative view might have recognized the need to factor in uncertainty (as opposed to risk) and encourage a 'plurality of thinking' as a means to accommodate it better.

The implication of embracing a complexity perspective on development aligns with a broadening of the current narrow understanding of IE as 'best' implemented through experimental studies. As Piccioto (this volume) says: *The new definition ignores uncertainty and complexity by conceiving of all policy and programme interventions as experiments and by reducing evaluation inquiry to a single question: did the intervention work as intended?* IE that is to be useful for resolving wicked problems such as social injustice, climate change, global epidemics and natural disasters must be able to contribute findings that help us understand how those intended to benefit are experiencing change and why interventions work, or do not, under which conditions.

Conclusion

We have argued that IE in development can contribute to development effectiveness in diverse ways to improve learning, accountability and influence policy and practice. If we agree that these purposes are important, then we need to be clearer about *whose* learning counts, *whose* accountability is improved and *whose* influence is strengthened. Furthermore, we have suggested that *how* IE contributes to these processes merits further debate: a debate that needs to transcend narrow methodological tussles and engage fully with others' viewpoints. We propose five areas that we believe to be promising starting points for that discussion: standards; rigour and relevance; power and politics; the nature of evidence; and complexity and uncertainty.

For this debate to be more productive requires a genuine desire to navigate with curiosity and listen across disciplinary codes. All of us need to be open to the possibility that our preferred approaches to IE are as likely to have limitations in terms of the purposes described above, as the projects they evaluate.

Notes

1. The report also noted the problematic quality of many 3ie studies, which highlights the difficulty of doing IE well.
2. Clemens and Demombynes (2013, p. 9) comment: 'IE methods in today's literature are better adapted to funders' questions than to implementers' questions. … . Pritchett *et al* discuss how iterative applications of IE methods, embedded within organizations in real time, can help implementers 'crawl the design space' to form and answer some questions that arise in the day-to-day process of making a project successful.
3. Implying that our particular 'disciplinary code' exerts some kind of agreement pressure is a bizarrely unconstructive tangent that is hardly central to the topic of this special issue.
4. Harrison (2014) confuses the process of accountability with what accountability reports on. The process, that is, nature of the relationship between commissioner and evaluator, often determines the focus – hence its importance.
5. http://en.wikipedia.org/wiki/Wicked_problem.

References

Barder, O. (2012) Complexity and development (presentation and podcast), http://www.owen.org/blog/5723.

Beinhocker, E. (2007) *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics*. Random House Business.

Big Push Forward (2013) http://bigpushforward.net/wp-content/uploads/2013/09/BPF-PoE-conference-report.pdf.

Booth, D. (2012) Development as a Collective Action Problem. Africa Power and Politics Programme Policy Brief 9,D.

Brown, D.L. and Jagadanda (2007) Civil society legitimacy and accountability: Issues and challenges. Hauser Centre and Civicus, http://www.civicus.org/new/media/LTA_ScopingPaper.pdf.

Brown, D.L. (2008) *Creating Credibility: Legitimacy and Accountability for Transnational Civil Society*. Kumarian Press, Sterling, Virginia, USA.

Claremont Graduate University (2006) Symposium '*What Constitutes Credible Evidence in Evaluation and Applied Research*'? http://www.cgu.edu/pages/4085.asp.

Clemens, M. and Demombynes, G. (2013) The New Transparency in Development Economics: Lessons from the Millennium Villages Controversy. Center for Global Development. Working Paper 342, http://international.cgdev.org/sites/default/files/Clemens-Demombynes-new-transparency_2.pdf.

Davidson, J. (2013) Understand causes of outcomes and impacts, http://betterevaluation.org/plan/understandcauses; http://betterevaluation.org/events/coffee_break_webinars_2013#webinarPart5.

Davies, W. and Haldane, A. (2012) Central banks should admit their mistakes: An interview with the Bank of England's Andy Haldane, http://www.opendemocracy.net/ourkingdom/william-davies/central-banks-should-admit-their-mistakes-says-bank-of-england%E2%80%99s-financial.

Deaton, A. (2013) New York Development Research Institute. Debate with A. Banerjee, http://www.nyudri.org/initiatives/deaton-v-banerjee/.

IFPRI (2013) http://www.ifpri.org/dataset/mexico-evaluation-progresa.

McGrory, E., Irvin, A. and Heise, L. (2009) *Research Rashomon: Lessons from the Cameroon pre-exposure prophylaxis trial site*. Washington DC: Global Campaign for Microbicides at PATH, http://www.global-campaign.org/clientfiles/Cameroon.pdf.

Morton, J., Shaxson, L. and Greenland, J. (2012) Process evaluation of the international initiative for impact evaluation (2008–2011), http://www.3ieimpact.org/media/filer/2013/01/07/3ie_proces_evaluation_2012_full_report_1.pdf.

Newman, K., Capillo, A., Famurewa, A., Nath, C. and Siyanbola, W. (2013) What is the evidence on evidence-informed policy making? Lessons from the international conference on evidence-informed policy making, INASP, Oxford, http://www.inasp.info/uploads/filer_public/2013/04/22/what_is_the_evidence_on_eipm.pdf.

Prewitt, K., Schwandt, T. and Straf, M. (eds.) (2012) *Using Science as Evidence in Public Policy*. National Academies Press, http://biblioteca.ucv.cl/site/colecciones/manuales_u/13460.pdf.

Pritchett, L., Samji, S. and Handammer, J. (2013) It's All about MeE: Using Structured Experiential Learning ('e') to Crawl the Design Space.Harvard University, John F. Kennedy School of Government. Working Paper Series rwp13-012, http://ideas.repec.org/p/ecl/harjfk/rwp13-012.html.

Ramalingam, B., Jones, H., Reba, T. and Young, J. (2008) Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. Overseas Development Institute, http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf.

Roche, C. (2013) Epistemology: The elephant in the room? http://bigpushforward.net/archives/2098.

Shutt, C. (2012) Transformational development and the politics of evidence? http://bigpushforward.net/archives/1688.

Woolcock, M. (2009) Toward a plurality of methods in project evaluation: A contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness* 1(1): 1–14.

