

Development is not an intervention: methodological issues in development evaluation

Fred Carden
Director, Evaluation
International Development Research Centre
fcarden@idrc.ca
www.idrc.ca/evaluation

Presented at the Virtual Conference on Methodology in Programme Evaluation (April 2010), coordinated by the Wits Programme Evaluation Group, University of the Witwatersrand, South Africa.¹

Introduction

*"The important tasks of investigating what works and what does not in the fight against poverty cannot be monopolized by one method."*²

Evaluating for development results opens questions:

- Whose development and whose results?
- Whose values are important?
- How do we define progress – development is continuous so we cannot define an end state but only continual improvement: here then we have to clarify:
 1. Improvement for whom?
 2. How do we avoid increasing inequalities?
 3. And, finally, knowing what happened is not enough: to improve, to use findings for policy purposes, to take ideas to scale, we need to know why, how, to the benefit of whom, in what context?

These are some of the questions I want to speak to here in discussing methodological issue we struggle with in evaluating for development results.

Let me declare my own position: development results are not about the project or program being implemented but about the change that is taking place on the ground. It is therefore essential to consider results from the perspective of what change is happening on the ground, not what change is happening in the program or the project.

¹ This paper was originally presented at the Sri Lanka Evaluation Association Conference Colombo, 22 April 2009, as "Evaluating for Development Results – Methodological Issues.

² Martin Ravallion. February 2009. Should the randomistas rule? Economists Voice. www.bepress.com/ev. p. 5.

Methodological plurality

“The main problem . . . is that they have put their preferred method ahead of the questions that emerge from our knowledge gaps.”³

I am going to offer a survey of the methodological issues that we confront as we try to improve evaluation and its contribution to development effectiveness. The papers that follow will be more concrete on some of the approaches that make up the range of possibilities in this rich and growing field.

Let me briefly make three points about method that I think are central to understanding what methods to use.

1. **First, form should follow function.** What this means is that the method of choice is never pre-determined but follows from what you are trying to find out, who needs to know and for what purpose they need to know. The above quoted critique is actually of random assignment as a method – but I removed the method identifier because this problem applies equally to any method where the proponent chooses first the method and then applies it to a problem. Many researchers are guilty of this and evaluation results suffer in consequence. For purposes of methodological development it is useful to have experts focused on a single method. For real world problem-solving however, method should be determined in context of the problem to be addressed.

2. **Second, values and political positions matter.**

The question then becomes, which or whose values and political stances are legitimately included in development evaluations? Which impacts get assessed?⁴

All social science, including evaluation, is conducted in contested environments where the science must dance with the values and politics of those who use the science. The science must contend with human volition and decision processes with all their uncertainties and indecision. So method cannot protect knowledge claims on its own and we should not allow ourselves to fall into this trap – it only leads to frustration and cynicism. What is most important here is to be clear on whose values and beliefs are included and whose are excluded. I will come back to this later.

³ Martin Ravallion. Ibid. p. 2-3.

⁴ Jennifer Greene, “What ‘evidence’ and what ‘methods’ can contribute to development effectiveness?” speech to the Perspectives on Impact Evaluation Conference, Cairo, March 2009.

In my estimation, we have the hardest-to-do science of them all! We do our science under conditions that physical scientists find intolerable. We face particular problems and must deal with local conditions that limit generalizations and theory building – problems that are different from those faced by the easier-to-do sciences. (p. 18)⁵

3. Third, Qualitative vs quantitative is a data question, not a methods question.

This is an important distinction. We often hear debates about the use of qualitative versus quantitative as if this were a methods issue. In reality most methods will make use of both qualitative and quantitative data. In the four standard social scientific approaches – experimental, statistical, case study and ethnographic – both types of data are used. Some will have more emphasis on one data type over another, but method itself does not limit data to only quantitative or qualitative.

Dani Rodrik describes the new mindsets emerging from these considerations as a shift from a traditional approach to policy framing to a new policy mindset:⁶

Traditional Policy Framing	New Policy Mindset
<ul style="list-style-type: none"> ▪ Presumptive 	<ul style="list-style-type: none"> ▪ Diagnostic
<ul style="list-style-type: none"> ▪ Long list of reforms 	<ul style="list-style-type: none"> ▪ Experimentation with lots of Monitoring & Evaluation
<ul style="list-style-type: none"> ▪ Complementarity among reforms 	<ul style="list-style-type: none"> ▪ Selective, narrowly targeted reforms
<ul style="list-style-type: none"> ▪ Best practices, rules of thumb 	<ul style="list-style-type: none"> ▪ Policy innovations
<ul style="list-style-type: none"> ▪ Straight mapping from policies to outcomes: testing innovations 	<ul style="list-style-type: none"> ▪ Experimentalist: innovation through implementation

Source: Dani Rodrik,

Building this shift calls for changes in how we think about methods for measuring development results. To get to this we need to **rethink** evaluation for development to shift our focus and priority away from the project or program and its funding to development effects on the ground. The political agenda has already moved here with the Paris Declaration. Practice, not only but especially in evaluation, is lagging. We need to **reshape** evaluation to take the local setting not the project or program as its unit of analysis. And we need to **reform**

⁵ David Berliner, 2002, quoted in Jennifer Greene: Speech to the Perspectives on Impact Evaluation Conference, Cairo, March 2009, page 4.

⁶ Dani Rodrik, The New Development Economics: We shall experiment, but how shall we learn? Paper for the Brookings Development Conference, May 29-30, 2008.

evaluation practice to directly address the asymmetries and inequities in North-South dialogue.⁷

This has methodological implications not only value implications. I have alluded to the value implications already. Let me elaborate the methodological implications because these are the main focus for this discussion. I will elaborate them around the agenda of **rethinking, reshaping** and **reforming** development evaluation proposed above. I hope that discussion will elicit some thinking on what you can do as evaluators and what commissioning and funding agencies (be they national governments or international donors) can do to take on this agenda.

1. Rethink:

“No method is sufficient to provide conclusive proof”⁸

Evaluation emerges from a tradition of examining discrete interventions – projects, change in academic method or medicine, and so on. Increasingly what we are concerned about in development is system change. This suggests we have to stop thinking about development as an intervention and think of it as a process where constant adjustment and experimentation is the norm. Evaluation must then find ways to be a useful tool in that more emergent process.

Rethinking development evaluation methods is defined by 3 key elements

1. Purpose of evaluation
2. Rigour: Evaluating evaluation (evaluation can be a negative)
3. Systems Orientation.

Purpose: Evaluation is most useful when it is built around the needs of the user, not the needs of the donor. We should go back to first principles about why development is initiated: development activities are meant to improve conditions in a community or in a society. It is therefore most important to think about the success, not in project terms, but in how change happens in society. Evaluation should be used to improve development not only to account for the expenditure of development funds. A use focus is essential.

Rigour: Any method can be more or less rigorously applied. This seems self evident but is often in dispute. As Ernest House points out in his recent article in the American Journal of Evaluation⁹ many evaluations are poorly done, including

⁷ Here I am drawing heavily on a collaborative paper, prepared by David Bonbright, Fred Carden, Sarah Earl, Sanjeev Khagram, Nancy MacPherson, Zenda Ofir and Patricia Rogers: “Impact Evaluation for Improving Development (IE4ID)”. It was first presented to the Impact Evaluation Conference in Cairo in March 2009. A short version is available at: http://www.idrc.ca/en/ev-142698-201-1-DO_TOPIC.html. The full paper will be published in 2010.

⁸ Ernest R. House. Blowback: Consequences of Evaluation for Evaluation. AJE, 29:4. p. 425. This is also reflective of Popper's notion of falsification as the only thing we can truly do: we can disprove, hence try alternatives, but we can never fully prove.

⁹ Ernest R. House. *ibid.* pp 416-26.

many scientific randomized trials. He points to numerous examples of sponsors having too much say in studies, about poor quality data, about opportunistic analysis of that data, about manipulation in testing and about control of authorship. Further both words and numbers can be manipulated or carefully used. Neither numbers nor words are absolute. The assumption that one method is by definition better and more rigorous is a dangerous one that leads to poor evaluation and results which cannot be trusted. The issue here is that rigour applies equally to all methods that we use and each must define and watch for its own rigour. That rigour can be tested using the criteria indicated above: Data quality, Methodological application, Data analysis methods, Control of authorship; and most importantly:

Comparison

Rigorous comparative case analysis (which can be done with a single case study or with multiple case studies utilizing different types of rigorous methods) is the primary strategy by which causal mechanisms can be identified, examined, tested, discarded and/or refined.¹⁰

We need to shift thinking to the higher order principles such as this. Baselines and counterfactuals – which are often demanded as essential to evaluation are not *a priori* truths¹¹. Rather, they are a manifestation of the higher order principle of comparison. Think about the purpose of a baseline or a counterfactual. Both are intended to give you something to compare with: in the absence of the intervention what would have happened; or at the start of the intervention, what was the condition? In both cases, we are talking about comparisons; but there is no reason to suggest that these are the only two modes of comparison that are universally valid. The core is that comparison is central. One good way of generating comparisons, especially in complicated and complex environments, is the case method. It is especially strong because it is designed to incorporate context not exclude it. The comparative case method generates the comparison and is equally valid to baselines or counterfactuals. The comparative case method may involve the use of one or multiple cases, but should always involve various types of comparisons – multiple methods and multiple comparisons is the watch word here.¹²

A key component of comparison is triangulation. Triangulation permits comparison across methods and across data sets. It helps to ensure data validity, construct validity, internal as well as external validity. Triangulation is

¹⁰ Sanjeev Khagram. n.d. paper on comparative case method. p. 1.

¹¹ Fred Carden. Baselines and Counterfactuals : Artifacts, not *a priori* truths. Posting to RealWorldEvaluation. May 2008.

¹² See Khagram, s & C.J. Thomas. Forthcoming. "Towards a Platinum Standard for Evidence-Based Assessment by 2020," in Public Administration Review.

important because reliance on a single data set and single method are not reliable.

Systems orientation:

“The importance of context for analysis of a phenomenon contrasts with experimental and statistical strategies that emphasize isolation of a phenomenon from its context in order to control and/or limit confounding variables”¹³

The final element of the rethink I want to raise here is context. Development happens in society so I argue that we have to integrate methods that embrace context and the system in which action occurs, rather than focus on methods that exclude context.

There is an essential heterogeneity in the outcomes of development interventions: not everyone responds the same way or makes the same choices in any situation (nor are their situations actually the same as noted by Pawson¹⁴): the un-observables come into play. In sum, context matters and needs to be taken into account in learning from evaluation for improvement.

2. Reshape

Reshaping evaluation methods based on this rethink leads us in the following directions:

1. First, build a robust Monitoring & Evaluation system within which evaluations are conducted. One-off evaluations are not enough to build a learning system. One-off studies are extremely limited whatever their method. A single case, or a single random study give very limited information. Without something to compare to, one has a hard time thinking about application beyond the case already completed. A robust monitoring and evaluation system itself gives some comparative opportunity and creates the conditions for evaluative thinking.
2. Use evaluation, so include knowledge translation as part of evaluation. Evaluation’s real value is in its use to ascertain worth and to also know how to move forward. Not all evaluators agree on this latter point. Scriven is quite clear that the evaluator has a responsibility to assess but not a responsibility to use that assessment for change. I beg to differ on this point as I see myself and other evaluators as part of social change and as part of building the democratic process. So questions about the implications of an evaluation’s findings for social change are in my view

¹³ Sanjeev Khagram. n.d. Notes on comparative case method. p. 1

¹⁴ Pawson, Ray. 2006. Evidence-Based Policy. Sage.

- important. They should be a central part of the evaluation system in any organization that is concerned with development and social change.
3. Articulate a theory of change. The theory of change language is much in vogue. Again, think back to the higher order principle. In this case, the principle is that we need to know what was intended and what path to change was being implemented if we are going to effectively assess what was done. Whether you call it Theory of Change, Outcome Mapping, or something else, the principle that is important is articulating what you intend to do, how you intend to get there, and doing so on an ongoing basis to take account of changes in conditions as implementation is underway. This leads to the second important aspect of a theory of change, that it is adaptive.
 4. Take account of the intended and unintended, positive and negative consequences. Not all methods are good at identifying unintended consequences, but some methods are. These should be integrated into the assessment to ensure that unintended consequences – which can be positive or negative – are considered. Many theory-of-change based approaches are particularly good at supporting this effort.
 5. Investigate causal mechanisms. Causal mechanisms are about more than causal attribution which is often what is sought in evaluation. Attribution allows you to say that, in general, a specific intervention causes a certain change. Causal mechanisms on the other hand, recognize context is key and that it is critical to understand the differential impacts of an intervention or a program on different groups in society, in different settings and at different times.¹⁵ Causal mechanisms are intended to help understand why change happened, who was affected, in what ways and in what contexts. It is this that program implementation and policy groups need to understand and know more about. Understanding causal mechanisms helps you much more with scaling up interventions than does attribution which can only tell you that the intervention worked in general in a certain way in a certain place. Causal mechanisms also help elucidate inequalities rather than masking them, precisely because you are seeking out the differential effects on different groups in a society. In these ways you are gathering the essential information you need to adjust and apply the program in other settings.
 6. Identify appropriate comparisons and make use of triangulation. Baselines and counterfactuals are two common approaches to comparison. Baselines allow you to compare with your starting point: what changed and how much since the start of the intervention? Counterfactuals allow you to compare with what would have been without the intervention. Both contain assumptions and it is a mistake to assume

¹⁵ See Pawson, *ibid*.

they are the best or only ways to achieve comparison. They are inappropriate where:¹⁶

- a. Interventions are multi-faceted; In the uncontrolled setting of a community it is often difficult to disaggregate out the individual actions and activities that are part of the project's intervention;
- b. The ground is shifting; changes are often embedded not only in projects, programs and specific interventions, but are embedded in the fundamental structures of our systems and societies – this is often described as 'chaos' or 'turbulence' or 'complexity'; we can no longer understand and modify by simply looking at past behaviour and outcomes; rather we have to look at patterns and relationships, understand what others are doing and how the context has changed; then we have to figure out how these things affect our own decisions; this means we have to be prepared for continuous change over time, leaving a baseline largely irrelevant in many situations; and third,
- c. Human choice is not predictable; people possess the characteristic of volition and we do not adequately understand how and why people make the decisions they make; there is little evidence to suggest that people make logical decisions or even decisions that best serve their own personal interests. There is rather more evidence, as Pawson suggests, that people make choices based on the capacities they have at a given point in time, the constraints under which they are operating (which may be moral, financial, social, technical, spiritual), and usually make their decisions in consultation with others (family, co-workers, bosses, friends) – all of whom have their own capacities and operating constraints. All of this leads to decisions in contexts of contestation.¹⁷ So, a counterfactual where human volition is involved is highly problematic.

7. Ensure Methodological Rigour

Rigour extends to all stages of the evaluation process:

- a. What do we want to know: Method selection and design
- b. What do we need to collect: Data quality
- c. What do we conclude from the data: Data analysis
- d. How do we know: The issue is comparative and triangulation not Baselines and counterfactuals

8. Reshape accountability to focus first and foremost on the intended beneficiaries.

There is an increasing understanding of the importance of accountability to the beneficiaries a program. Increasingly it is recognized to be at least

¹⁶ Fred Carden. Baselines and Counterfactuals : Artifacts, not *a priori* truths. Posting to RealWorldEvaluation. May 2008.

¹⁷ Ray Pawson. Ibid.

as important as accountability to the decision makers and funders of an initiative. Knowing if an initiative is making a difference on the ground is what we are really accountable for. Clearly the appropriate controls on spending are important, but if in the end we are doing the wrong thing to meet our goals in the fight against poverty, then it is unimportant that we are doing it well. We must be accountable to beneficiaries and ensure their voices are heard.

3. Reform:

In a truly experimental approach to development (and here I mean experiment as an approach, not the experimental method), monitoring & evaluation play a key role: if you innovate as you go you need good systems to know how things are evolving so you can adjust and modify on an ongoing basis and still maintain a flow of data about progress.

In particular, methods should:

1. Address asymmetries in evaluation
 - Methods should address the specificities not only the generalities to avoid exacerbating inequalities
 - Methods should seek causal mechanisms as they pertain to different groups and the contexts in which they operate
 - Methods should ensure they address the use needs of the purported beneficiaries of a program, not only the use needs of the funding agent (be that an external agency or a government agency)

2. Strengthen evaluation systems and architecture
 - Methods should be applied within an evaluation system that addresses inequalities, considers the system in which an intervention is evaluated and builds appropriate linkages both within levels of a system and across systems.
 - The evaluation architecture should include both evaluation and evaluative thinking

3. Be appropriately resourced
 - Evaluation takes resources. Too often these are not provided or siphoned off to 'more important' needs of programs. M&E are not fully integrated as useful and relevant but continue to be seen as a burden. It is up to evaluators to change that perception.

4. Adopt a range of reporting techniques to enhance use
 - It is insufficient to focus on method. Use is paramount and so the study should be considered from the point of view of communication of findings, adaptation of findings to different users and should include sufficient follow-up resources to give every opportunity for use.

5. Increase the capacity to do evaluation. So it must be inclusive and engage key parties.

Conclusions

These changes require strong leadership, both at the top and in the evaluation community. Getting better at evaluation means getting better at using methods appropriate to the question at hand. But we can never forget that method alone does not protect knowledge claims. We must learn and track the values and socio-political interests of those who make decisions. Evaluation should contribute to change. So it is not only about evaluating single projects or single interventions, but about development evaluation. As a colleague from Mauritania noted, we must never forget that,

"it is not about your project, it is about my country".¹⁸

¹⁸ Oumoul Ba Tall, speaking at the European Evaluation Society, October 2009.