

IMPACT EVALUATION FOR IMPROVING DEVELOPMENT (ie4id)

This paper was written by David Bonbright, Fred Carden, Sarah Earl, Sanjeev Khagram, Nancy MacPherson, Zenda Ofir, and Patricia Rogers

INTRODUCTION - WHO WE ARE

The authors of this paper come from a variety of perspectives. We are scholars, practitioners, and commissioners of evaluation. Our entry points, thematic interests, disciplines, geographic locale, and experiences may differ but we share a fundamental belief that knowledge (whether obtained through research or evaluation) has the potential to contribute to positive social change. In this sense, we share common ground with the authors of the influential report, “When Will We Ever Learn” and others such as Michael Patton who wrote,

“I practice and write about evaluation because I believe that evaluative thinking can make more effective those who are deeply committed to and authentically engaged in making the world a better place. Evaluation, at its best, distinguishes what works from what doesn’t and helps separate effective change makers from resource wasters, boastful charlatans, incompetent meddlers, and corrupt self-servers. Through evaluation, I aspire to make my own small contribution toward realizing the vision of an experimenting global community, one characterized by commitment to reality-testing, respect for different perspectives, and open dialogue about evidence – a world in which ongoing learning is valued and practiced and knowledge is generated and used.”¹

We know that the full potential of evaluation is not always (or even often) realized in international development. There are many reasons for this – some to do with a lack of capacity, some methodological, some due to power imbalances, and some the result of prevailing incentive structures. Our position is that evaluation, like development, needs to be an open and dynamic enterprise. We think that some of the current trends in evaluation, and especially in impact evaluation in international development, are unhealthily narrowing the field. Through this article we are trying to keep space open and encourage others to join with us in refusing to concede:

- development and its assessment to technocrats;
- rigour to a single method; or,
- evaluation to one user or purpose.

¹ Patton, Michael Quinn. (2008). Utilization-Focused Evaluation. 4th edition. Thousand Oaks, CA: Sage. Pp. xvii-xviii

We believe that for evaluation to be useful, rigorous, feasible, and proper (by our professional standards of good quality and ethics), we need to demand openness and diversity. This article is part of our contribution to discussions about how to realize this potential and our commitment to work towards it. We hope these ideas will help legitimize and further others' words and actions. None of us are alone in our efforts.

EXECUTIVE SUMMARY

Impact evaluation can make a difference to development. Accurate, feasible and useful evaluation that is aligned with the type of development initiative it is meant to assess can help make development better – by guiding improvements in policies, programs, projects and practices, identifying and explaining successes (so they can be emulated) and failures (so they can be avoided), informing investment decisions, and encouraging funders and partners to maintain their support. There is, however, another side to impact evaluation. Low quality or misaligned impact evaluation can waste scarce resources, reinforce inequalities, and move us away from development lessons already learnt by supporting the adoption of ineffective or inappropriate policies or practices and undermining effective ones.

The promise of well defined and properly implemented evaluation, as well as the threat to development of poorly defined and implemented evaluation have brought us together. In this paper we set out an agenda for action to ensure impact evaluation fulfills its promise. We argue for a decisive move away from impact evaluation *OF* development to impact evaluation *FOR* development – impact evaluation that not only assesses development but consciously and demonstrably contributes to improving development. This paper sets out a seven-point agenda of rethinking, reshaping and reforming impact evaluation for improving development. We propose actions that can be taken by all those involved in impact evaluation of development - practitioners, thought leaders, agenda-setters, as well as managers and commissioners of evaluation.

1. Impact evaluation should contribute to improved development that can be sustained in the long term. Impact evaluation must ultimately be judged by its effect in terms of actually improving development. This means that technical merit is necessary but not sufficient. Timeliness, relevance, usefulness and the actual effects of basing development policies, programs and projects on evaluation must be considered. These need to be looked at from the points of view of the different constituents - especially primary stakeholders - to the development initiative being evaluated. Impact also has to be evaluated on whether it is likely to contribute to the capacities of local people institutions and systems in the longer term.

2. Impact evaluation should suit the nature of development. The selection of impact evaluation methods and processes should be appropriate to the context of development and aligned with the type of development initiative under consideration. Impact evaluation should not start with particular evaluation methods and approaches.

3. Impact evaluation should draw on the full range of methods and designs for systematic and rigorous empirical investigation, including ethnographic, case study, statistical and experimental/quasi-experimental approaches and methods, as well as those from such fields as complexity science, participatory research, and action learning. No one approach is inherently more rigorous than another. Rigor is dependent upon both *appropriate* method choice – selecting methods based on evaluative purposes and contexts – and successful implementation that meet accepted standards of merit.

4. Impact evaluation should produce a comprehensive analysis of impacts in terms of different outcomes of value (including intended, unintended, positive and negative), as well as the distribution of results, costs and benefits. How local actions are affected by national and international systems, strategies and policies and vice versa should be assessed.

5. Impact evaluation should explain how and why impacts occur. A good understanding of how results have been achieved (what works, under what conditions, by whom, how and why and for how long) is essential for learning and improving from successes and failures. In practice, this means that knowledge about average impacts is insufficient to inform decision-making; at the very least rigorous examination of causal mechanisms and contextual factors is required when transferring initiatives to new settings and scaling is of interest.

6. Impact evaluation should be embedded within robust systems of monitoring, assessment and learning. Isolated, discrete impact evaluations are of less value and less useful than impact evaluations that are part of a robust system of monitoring, evaluation and learning. Embedded impact evaluation creates synergies between real time adaptation and improvements with periodic assessment of results to increase accountability. Impact evaluation is one element of the process of understanding what works, under what conditions, by whom, how and why and for how long. Integrating impact evaluation within a robust monitoring and learning system strengthens its contribution to improving development - it should not unnecessarily divert resources away from other useful types of evaluation.

7. Rethinking and reshaping impact evaluation requires fundamental reform. Immediate steps can be taken to practice appropriate evaluation that is focused on improving development and to reverse the tendency for impact evaluation that does not. The distribution of power, prevailing role definitions and incentives in development often push toward bad evaluation. A deliberate reform effort is essential. The recommended changes in the practices of key actors in development are meant to reinforce each other so that the overall system of development is improved through better evaluation.

RETHINKING IMPACT EVALUATION

Starting from current practice, there must be two cardinal principles about impact evaluation that rise above others: it contributes to improved development and it suits the nature of development.

1. MAKE A DIFFERENCE

Impact evaluation for improving development (ie4id) needs to be rethought to respond to significant changes in what 21st century development entails and how we understand it. In the past few decades a

range of development related practitioners and scholars have generated new insights into what constitutes development, and along with that an expanded view of what constitutes development effectiveness.

Development has changed significantly in the past few decades due in part, though not exclusively, to processes of globalization. In particular, 21st century development initiatives are now more often complicated, and even complex, in the following ways:

- Multiple goals of development associated with a wide range of issues are simultaneously being promoted
- Development effectiveness involves a wider range of dimensions
- Broader sets of legitimate stakeholders (beyond aid agencies and national governments), along with a range of sources and types of capital are actively engaged
- Primary constituents are recognized as central agents of development
- Contextual factors are understood as critical conditions that shape development
- Diverse initiatives beyond discrete, standardized projects are being implemented
- This broader array of initiatives often involve multiple interacting or alternative causal pathways, or non-linearity, recursive causality and emergence
- Effective development often requires experimentation, iteration, learning, and adaptation in real time.

Development clearly involves more than donors and national governments. Yet, evaluation practice often tends to privilege them as the primary audience of evaluations generally and impact evaluation specifically. Citizens, civil society, philanthropic organizations, the private sector, regional and global organizations, partnerships and networks all contribute to development. In particular, there is increasing recognition of the role of community members as primary agents of development, not as passive recipients of aid. Impact evaluation therefore needs to find practical, yet meaningful ways to engage the range of actors in the process and plan to address their specific information needs.

Impact evaluation should increase and improve accountability not only to funders and decision makers, but also to primary constituents and other key stakeholders. The outcome of mutual accountability is deeper legitimacy and improved effectiveness. Narrow legitimacy is achieved through accountability towards funders and other partners. A more profound legitimacy is achieved especially when those who are meant to benefit from development are able to hold those who promote development to account.

Too many development efforts fail because of an insufficient focus on creating those conditions that give the best chance for long-term success. Development initiatives are not only about producing results in the short-term, but also about developing individual, institutional and system capacities for ongoing results, thus enhancing the chance that development will be sustained. Impact evaluation needs to contribute to both of these. It can inform short-term decisions about policies and programs. It can also help to build capacity for improved planning, implementation and learning towards long-term results. Rethinking the role of impact evaluation includes considering how it can support or undermine self-

determination, constructive negotiation, and improved results in the long term – and following up to see if this is actually the case.

2. WAG THE TAIL

Many of the methods and approaches used today in impact evaluation have been developed for discrete, standardized interventions, like drugs or seed varieties, where the intended results, and how to measure these, are clear and agreed, and controlled comparisons of with and without are both possible and meaningful. For development interventions like the above, these methods can work well when they are adequately resourced and well implemented. Even in these cases, impact evaluations can be made even more rigorous and useful through the utilization of mixed methods, and by embedding them in robust systems of monitoring, assessment, learning and improvement.

But many development initiatives are not like this. They are broadly defined approaches or strategies, often implemented in different ways in different situations, and obviously influenced by varying local forces. Simple with/without comparisons are often not the sole or most meaningful means of assessing impacts because the initiatives (and their underlying change logics) have complicated aspects (multiple components that vary in different settings), complex aspects (dynamic and emergent), or both.

This has particular implications with the increasing emphasis on drawing on the evidence base to inform development practice and funding decisions. If the only evidence that is seen as credible is the type of evidence that can only be generated on discrete, standardized interventions, then there is a risk that multi-faceted or systems-based initiatives will not be supported – even if they are in fact effective. This will have dire consequences for development.

So what are we to do? Avoiding these realities is not an option. Failing to undertake rigorous impact evaluation of initiatives with complicated and complex aspects, and relying on their face validity, or good intentions, or selective descriptions of direct beneficiaries, is demonstrably not sufficient. But neither is it appropriate to mould these interventions into discrete, standardized programs to make them easier to evaluate – that would be like the tail wagging the dog. Impact evaluation for development has to be able to evaluate in a rigorous manner the impact of complicated and complex initiatives, and provide useful advice to support the translation of findings to other situations. And it can, by following the principles for reshaping and reforming impact evaluation below.

RESHAPING IMPACT EVALUATION

To enact these cardinal principles, the practice of impact evaluation should be more scientific, avoid a narrow focus, recognize differences and close the loop.

3. BE MORE SCIENTIFIC

We applaud the increased emphasis on more scientific impact evaluation (with scientific being one element of rigor). Discussions around “what counts as credible evidence?” and “what makes evaluation rigorous?” all too often digress into debates focused on selection of methods and research designs. Efforts to improve the technical merit of impact evaluation have defaulted unnecessarily to a narrowly

defined set and inappropriate hierarchy of methods. This narrow definition of methodological rigor is driven by conceptualizations of science that are inaccurate and outdated.

Impact evaluation must be authentically scientific in data and method selection, which depends on the type of initiative, chosen evaluative purposes and context(s) at hand. In other words, scientific rigor should be judged in terms of appropriate data, method selection and implementation. Data and method should follow purpose.

This definition of methodological rigor is based on insights and techniques that have been developed across natural, social, systems and complexity sciences. It identifies and addresses threats to internal and external validity that are present in any measure, method or research design. Validity also has an inherently cultural dimension – all data, methods and analysis must address the challenge and opportunity of multiple cultural understandings of reality.

Scientific impact evaluation emphasizes multiple kinds of valid comparisons, including the use of counterfactuals when appropriate, triangulation across methods and types of quantitative and qualitative data to improve measures and analysis. Sophisticated diagnostic tools for appropriate (and often mixed) methods selection and measurement are increasingly available and should be utilized, as are and should appropriate norms, guidelines and criteria to provide assurance that any method or set of methods meet technical and ethical standards.

4. AVOID A NARROW FOCUS

Development has many dimensions and many development initiatives in the 21st century have multiple goals. While it is not possible for every impact evaluation to cover all possible aspects of a development initiative, impact evaluation should take into account the purposes, context and system surrounding a project, program or policy. It should produce an adequately comprehensive assessment of results, as well as sufficient information about context and processes, for the findings to be potentially applicable to other settings.

There is especially an ethical imperative to identify and analyze critical unintended consequences, especially negative ones. Impact evaluation should deliver a balanced assessment that includes intended and unintended, positive and negative impacts; assesses the distribution of benefits and costs of initiatives; addresses effectiveness and also legitimacy, efficiency and sustainability; and evaluates beyond the boundaries of the initiative to determine how global systems, strategies and policies actually affect local actions and vice versa.

5. RECOGNIZE DIFFERENCES - UNDERSTAND UNDER WHAT CONDITIONS, FOR WHOM, AND IN WHAT WAYS DEVELOPMENT WORKS

Improvement in impact evaluation means building knowledge about what works, under what conditions, by whom, how and why; assembling knowledge about when and in what ways initiatives should be transferred to other settings; and how to generate and use evidence to inform policy and practice. When impact evaluation does not recognize local knowledge, cultures or situational differences in results, then it tends to impose 'one-size-fits-all' policies to the detriment of improving development.

Data about the average intended effects of a project is critical but often not sufficient to inform good policy or practice. Information is also needed about the results for different groups (particularly for the most disadvantaged) to inform judgments about the value, transferability and scalability of a development initiative. Information is also needed about the contexts in which the project is successful to inform conclusions about the generalizability of the findings. Few development initiatives are expected to work the same in all cultures and settings at all times. The task is to bring understanding to how it is and may be affected by local and global contexts. Understanding impact heterogeneity is at the heart of good impact evaluation.

A range of designs can be used to discover and describe different impacts. It is possible for experimental and quasi-experimental methods for impact evaluation to pay attention to impact heterogeneity if relevant variables can be identified and measured in advance. However, it is frequently difficult, if not impossible, to anticipate (and account for) many relevant factors. Impact evaluation approaches that draw on a wider range of credible evidence, especially multiple kinds of comparison and triangulation, make it possible to include variables that may not have been identified or anticipated in advance.

For some initiatives it will be possible to identify in advance how pre-planned activities and outputs could produce clearly defined results. For many development initiatives, however, no single actor, factor or causal mechanism working independently will be sufficient to produce results. Rather, impact could be attained through the combined efforts of multiple actors, with overlapping and interacting causal mechanisms, in conjunction with particular contextual factors. Solutions will involve changing beliefs, attitudes, relationships, capabilities, conditions and behaviors and often need to be worked out over time with constituents as well as a range of partners. For these types of initiatives, an iterative and adaptive approach is essential and impact evaluation will need to document and support this.

6. CLOSE THE LOOP – EMBED IMPACT EVALUATION

Producing rigorous evaluation results at periodic intervals is not sufficient. Good impact evaluations facilitate utilization through processes and structures that close the loop back to decisions and actions by those for whom the evaluation was commissioned and who committed to using it to guide their decisions or influence their actions. Impact evaluation is only one type of evaluation that can contribute to improved development. Resources for the former should not detract from the later.

Impact evaluation is more effective when it is part of a robust learning and improvement system that integrates across operational cycles of planning, monitoring, assessing, reporting, and adapting. Such learning and improvement-oriented systems translate credible evidence of a development process or initiative into actionable guidance, especially for policy makers and citizens. They do so in real time in ways appropriate for effective utilization. Embedded impact evaluation creates synergies between real time improvements and longer-term assessment of results.

The effectiveness of development initiatives will especially improve when they are held accountable by those most affected - positively and negatively. This implies that public reporting of impact should reflect the ways in which a development initiative responds to and balances the needs and perspectives of its various constituents. Impact evaluation design needs to explicitly consider: (a) how the initiative translates its understanding of what is happening into action; (b) how it adapts and improves; and (c) how different constituents of an initiative understand and act on evidence about results.

REFORMING IMPACT EVALUATION

To create a more enabling environment for impact evaluation, each of us - as evaluation practitioners, commissioners and managers of evaluation, thought leaders, or agenda setters, educators and mentors - have critical contributions to make. Fundamental reform requires each of us to ask, "How can we do better?" and "What do we need to do differently?" to implement these changes.

7. CHANGE THE SYSTEM

While we recognize the importance of power and politics in improving ie4id, we also argue that everyone involved can take action and contribute to positive change. We have identified below an initial list of steps to reform the system.

What should evaluation practitioners do?

Evaluation practitioners are critical actors in this system. They are the implementers of impact evaluation policies and approaches on the ground. Practitioners, particularly practitioners who work in their home countries, have position, power, knowledge and access to the projects, programs and policies where development action is implemented. While they do not normally independently carry the mandate to effect institutional change, there are many steps evaluators can take to move us towards a vibrant field of ie4id.

1. Challenge yourself; challenge your clients.
 - a. Be a model of good practice; maintain high quality using appropriate evaluation standards; ensure that you follow impact evaluation practices across disciplinary boundaries.
 - b. Recognize that your clients may not be clear on what they need, what degree of certainty evaluation can provide and what can realistically be achieved within time and resource constraints. Help them to identify their needs and make decisions in terms of use-oriented impact evaluation.
 - c. Promote impact evaluation from a systems perspective. Understand your assignment in the wider development context not solely in the narrow parameters of the initiative.
2. Promote appropriately rigorous methods.
 - a. Be a learner as well as innovator: do not become complacent in your practice but continue to monitor changes in the environment follow and participate in debates, seek out and contribute to new tools and methods and follow and contribute to standards for good impact evaluation practices.
 - b. Recognize that your values and practices are shaped by your history, as are those of others; doing impact evaluations in other cultural contexts is therefore extremely sensitive and benefits from active dialogue/participation with key stakeholders.
 - c. As an educator, open your students up to multiple approaches and methods in impact evaluation, drawing from different disciplines, perspectives and worldviews.
3. Build vibrant evaluation networks.
 - a. Seek out and work with others through associations and communities of practice. These provide not only professional support but are also key mechanisms for reforming the

policies and agencies that shape impact evaluation practice. Do not settle for passive engagement.

4. Contribute to improved reporting and knowledge sharing.
 - a. Focus on effective, tailor-made reporting and communication of impact evaluation results, methods, tools, advice, lessons and examples, to reach a broader range of audiences across disciplines, practice, policy and research.

What should the commissioners and managers of evaluation do?

Those who commission and manage impact evaluations are in an extraordinary position of privilege. Whether they recognize it or not, they have the power and the authority to positively change the way impact evaluation for improved development is conceived of, commissioned, managed, reported and disseminated. Conversely, by failing to recognize, or recognizing but not acting on, the prevailing inequities and biases of much current development evaluation practice – including impact evaluation, they perpetuate inappropriate practices. Those who commission and manage evaluations must be strong, independent thinkers, well grounded in development as well as evaluation theory and practice, transparent about values and practices and brave enough to invite and listen to those who will ‘speak truth to power’.

Evaluation commissioners and managers must themselves be committed to learn from success, but also be prepared to face up to the failures of development and do the necessary work of shattering ill-founded dogmas, questioning self-serving assumptions and challenging complacency among all stakeholders. While a tall order in many political environments and organizations where ‘safe spaces’ for evaluation and learning are limited, there are practical steps that all commissioners and managers can take, to varying degrees, to create a supportive authorizing environment and set of incentives towards ie4id:

1. Be clear how impact evaluation is intended to contribute to improving development
 - a. Conceptualize and frame evaluation around explicit theories and assumptions of what contributes to effective development and improvements in the lives of people and the systems upon which they depend (use explicit definitions of development and hypotheses about the role that rights, choices, freedom, economic growth, sustainability, accountability, etc. play in development processes).
 - b. Focus on outcomes and impacts to the extent feasible and appropriate, not only outputs.
2. Set, promote, use and support the improvement of standards and guidelines for development evaluation that reflect international good practice and regionally adapted standards (such as the African Evaluation Guidelines).
3. Commission evaluations that push the boundaries of existing approaches to impact evaluation and that evaluate beyond the interests of those in charge of specific development initiatives.
4. Join forces with other organizations and agencies to broaden the objects of what is evaluated to capture what matters in development, for example, the drivers and root causes of un-sustainability; the effects of developed country policies on developing countries in areas such as foreign investment, trade, globalization, migration, pollution, intellectual property, private capital flows;

and the importance of evaluating capacity development at individual, institutional and systems levels as a critical step towards the achievement of development outcomes and impact.

5. Embed evaluation – and hence impact evaluation - within the management and leadership of organizations. Make sure evaluations consider the way an organization or initiative plans, monitors, assesses, reports back to constituents about results, and reports out to the public.
6. Untie funds for impact evaluation for development
 - a. Manage the procurement process to prioritize the use and development of high quality regional, national and local evaluation skills.
 - b. Set up pooled national funds at arm’s length from power-holders to enable less empowered if not marginalized groups and organizations to engage their own evaluators to assess development initiatives.
7. Invest in capacity and field building in impact evaluation at multiple levels – developing individual skills, organizational infrastructure, networks of practitioners and professional evaluation associations.
8. Commission and support improved reporting and communication of evaluation results, methods, tools, advice, lessons, and examples, in order to reach a broader range of audiences across disciplines, practice, policy, and research. This could include authoritative, user-friendly briefs on progressive approaches to and examples of impact evaluation to balance the body of information available on impact evaluation.

What should thought leaders, agenda setters, educators, and mentors do?

Another powerful group is those who have the responsibility and opportunity to create impact evaluation knowledge and influence, teach and mentor the impact evaluation commissioners and practitioners of today and tomorrow. As the field of evaluation - including impact evaluation - becomes more reflective, it calls for champions who can communicate its value, who can be influencers in the halls of power and motivate and engage the broader public. They have to bring theory to practice and improve theory based on good practice - informing new ideas, developing new concepts, and challenging conventional wisdom.

Ultimately we all follow their innovations. It is therefore extremely important that they move beyond narrow discipline-based confines towards broad perspectives on impact evaluation in line with the principles set out in this paper. As development is increasingly recognized as trans-disciplinary, so too should training and capacity building in impact evaluation bring together the fields of social, natural and emerging sciences, management and cultural disciplines to provide much needed insights into complex questions of behavioral, institutional and structural change. Thought leaders, agenda setters, trainers and mentors should therefore:

1. Articulate and challenge ill-founded dogmas about development and evaluation, question self-serving assumptions, and challenge complacency among policy makers in both evaluation and development.
2. Provide ‘safe spaces’ for evaluative learning, analysis and reflection within academic and research environments, and in civil society and government.

3. Bring together citizens, civil society, business and public sector to examine new ways of collaborating on impact evaluation.
4. Incorporate accountability frameworks, transparency requirements, ethics, independence, inclusive methodologies, and citizen engagement in new ways of practicing development evaluation and impact evaluation.
5. Promote improved reporting and knowledge sharing of evaluation results, methods, tools, advice, case studies, and examples, to reach a broader range of audiences across disciplines, practice, policy, and research.

Final Thoughts

We are privileged to work in a field where our findings and processes can change lives for the better. Our field is expanding. It is currently in the spotlight amongst politicians and policy makers. Fundamental changes are both necessary and possible. This paper points a way to making impact evaluation more relevant, credible and useful for improved development.

To those who want to make a difference, to those who want to bring about change, we extend an invitation to help us strengthen this call to action. There is much to do, let's get to work.