

Evaluating for Development: Reshaping our evaluation criteria?

“... [G]rand overarching theses that encompass the whole of the development paradigm are unlikely to thrive, and development thinkers of tomorrow will need a new humility when faced with the growing complexity of the development terrain”. Curry-Alder et al, 2014.

This paper argues that “development evaluation” does not attend sufficiently to “development’, especially considering the challenges facing the Global South. It is necessary to shift to a more dynamic approach that will enable us to evaluate FOR development. This, in turn, will compel us to make sure that the criteria that determine “what” and “how” we evaluate are refined in concept and in practice, cognisant of what defines an intervention that truly fosters development.

Zenda Ofir, Evaluation Specialist

Introduction

THE development evaluation community is aware that our evaluation criteria, especially the ubiquitous OECD Development Assistance Committee (DAC) criteria, have to be updated. Thoughtful contributions have argued for refinements (e.g. *Picciotto, 2014*) or real shifts (e.g. *Heider, 2017*) in concept and practice. But what will shape them in future? On what basis do we decide which criteria to set aside, refine or add?

While “development” is now according to the 2030 Agenda the charge of all countries, the challenges remain the most severe among low-income countries. This paper argues that “development evaluation” does not attend sufficiently to “development”, especially considering the challenges facing the Global South. It is necessary to shift to a more dynamic approach that will enable us to evaluate FOR development.

This, in turn, will compel us to make sure that the criteria that determine “what” and “how” we evaluate are refined in principle and in practice, cognisant of what defines an intervention¹ that truly fosters development.

The paper aims to stimulate debate on this issue by proposing several characteristics of “evaluation for development” (E4D) that should help us to reshape our evaluation criteria, and the way they are applied to

interventions – while acknowledging that “interventions” are not the only focus for evaluation, but an important one. It also highlights the interesting similarity between the holistic approach to health interventions found in traditional oriental medicine, and the notion of evaluation for development viewed through a complex systems lens.

A new era for development evaluation?

Over the past two decades the evaluation profession has become truly global, and development evaluation² has been thriving. We have now entered an era where the challenges confronting evaluators working in a development context urgently demand new ways of thinking and working (*Picciotto, 2015*). Development blueprints, dominant ideologies and notions of “best practice” are being replaced by a colourful diversity of frameworks, models, goals and practices. At the same time, natural and man-made crises and disasters are destabilising large parts of the world, while countries and regions jostling for power are influencing development potential and strategies. The private sector is set to become a much more active and visible funder of development. And we are only now beginning to understand how development will be influenced by a hyper-connected world where the ambition of the 2030 Agenda with its Sustainable Development Goals (SDGs) is set to intersect with the Fourth Industrial ►



► Revolution,³ and the digital economy with the “human economy”.⁴

The emphasis in the 2030 Agenda on the fact that development is the charge of all nations highlights the fact that development is a path, not a destination. This presents a good argument for merging “evaluation” in the Global North and “development evaluation” in the Global South.⁵ Yet development in Switzerland, Japan or Canada means something very different from development in Nepal, Yemen, Mali or Paraguay. Tweaking policies or sectors amidst robust institutions and systems while working from a strong foundation (based on development indicators) may be demanding. But it is clearly far less demanding than establishing a positive development trajectory from a weak foundation, over many fronts, for a prolonged period, while power, resource and capacity asymmetries interfere.

This means that we can blend “development evaluation” completely with “evaluation” only if we are sure that our theories and practices truly enable us to serve those countries and societies that face major development challenges. After all, the foundational theories that shaped our profession originated in rich Western countries with little reference to the contexts that define the Global South. We therefore have to make sure that our evaluation theories, practices, standards, criteria and questions help us to evaluate for development that genuinely improves the wellbeing of all our societies, and their ecosystems.

The problem: Development evaluation does not necessarily foster development

The point of a development intervention is that it should contribute to development.

But how often do we consider exactly what is meant by that? Have we been too laid-back in our engagement with the concept of development? Have we become overly complacent consumers of statements about development effectiveness and development evaluation?

Development remains a contested concept; the differences between bottom-up/top-down, or micro/macro perspectives continue to lead to vastly different ideas and ideologies about how it can be achieved (*Curry-Alder, 2014*). As a result, it is very difficult for an evaluator to judge the merit of a specific development model during a certain phase in the evolution of a country or region. But we can be explicit about the ideology, values and model that we use when we evaluate the merit, worth or significance of a development intervention. This is displayed, in part, by the evaluation criteria that we choose to focus our interventions, and by how we apply them in practice.

In doing so, we have to consider the following:

- Development progress is almost always considered and measured at country (i.e., societal, national) and regional level, as the proliferation of global indices and the reporting modalities for achievement of the SDGs illustrate.
- Definitions of development effectiveness confirm the need for collective, if not coordinated, action by a range of local and external actors. Development effectiveness is a measure of these actors’ aggregate impact, and accountability for results is shared by many.
- From a national perspective, we can get “development without development” (*Chang, 2010*). This happens for example when so-called development ►



- ▶ interventions focus on enabling conditions such as poverty reduction, individual betterment or meeting basic needs without a vision of how the country can sustain a positive development trajectory in the long term.
- A country that shows development progress will need to have positive trajectories across multiple interconnected aspects of the life of its citizens. A narrowly focused intervention that does not at the right time unfold ▶▶

- ▶ in concert with the other interventions or goals, or that leads to localized or insignificant outcomes and impacts compared to what has to be achieved to ensure a positive development trajectory, cannot be assessed as a meaningful contribution to development.
- This is especially problematic in what Chang calls “ersatz development” – whether supported by aid or philanthropic agencies, the private sector or a government – that relies on uncoordinated interventions that do not build on synergies between interventions, and do not enable system coherence that can facilitate change at a more macro level.
- If achieved outcomes or impacts do not sustain, or foster spontaneous action that leads to further outcomes and impact ripples, a so-called development intervention cannot be said to be successful. It can even lead to regression if people become disillusioned and disheartened as a result.

Thus, not every “development intervention” in a low-income country contributes to development, even if it is part of a national or local development plan, meets “community” expectations, achieves desired outcomes or impacts, and/or shows a “silver bullet” effect. Or even if it successfully strengthens capacity, builds infrastructure, enhances access to health services, develops leadership or saves lives.

Not only do we often fail to attend to the “higher levels” of theories of change that supposedly connect an intervention with a specific development context or model, but also our evaluation criteria do not necessarily compel us to attend in practice to all the critical issues that matter for development.

The solution: Evaluate for development

Placing a focus on the notion of a more dynamic, holistic approach to evaluating for development might help us to refine our criteria. This approach to evaluation has as a main premise that we have to fully recognise the implications in practice of viewing development as a complex adaptive system (CAS) (Ramalingam, 2014). We then need to make sure that we embed this understanding in our evaluation criteria.

From this perspective, evaluation for development (E4D) has at least five relevant characteristics⁶:

“Not only do we often fail to attend to the ‘higher levels’ of theories of change that supposedly connect an intervention with a specific development context or model, but also our evaluation criteria do not necessarily compel us to attend in practice to all the critical issues that matter for development”.

1. Development as CAS

E4D views development through a complex adaptive “socio-ecological” systems lens.

E4D acknowledges that development, and development interventions, are complex adaptive social-ecological⁷ systems (Ramalingam, 2014; Orstrom, 2009). This means that evaluators have to grapple with the implications of dynamics and properties such as interdependence, non-linearity, coevolution, path-dependence, self-organisation and emergence that affect the relationships between ▶

► and within interventions, and the resulting patterns in society. We will need to sharpen our evaluation foci and criteria, bearing in mind that a complex adaptive systems approach has implications for evaluation practice far beyond evaluation-informed adaptive learning and management.

2. Attention to preconditions

E4D shifts attention from overly simplistic theories of change and rigid notions of “results” to a stronger focus on preconditions for success.

The current wave of “theories of change” based planning serves several important purposes. Yet they are often so simplistic compared to reality, and the underlying assumptions and results are sometimes so poorly conceptualised, that they become almost meaningless. Both development planners and evaluators need a stronger focus on identifying and understanding possible preconditions⁸ for success. This is with a view to improving the likelihood that the positive development outcomes and impacts will be achieved, and sustained, within an evolving context.

Preconditions can, to some extent, be identified through existing knowledge and insights (gleaned from the literature and experience), understanding the disposition of a society in a specific context, and evidence-informed learning.⁹

3. Attention to trajectories

E4D shifts attention from results as snapshots, to intervention and development trajectories.

The current emphasis on adaptive management draws attention to the need to make assessments cognisant of trajectories (at different spatial and temporal

scales). For example, the path followed by an intervention to contribute to desired development impacts, which in turn relates to the development path of a society, country or region.

An intervention might stall or fail before yielding results. It can suddenly reach a tipping point or fail due to a change in context that blocks an essential pathway. Thus, if evaluation is done at the wrong moment, the findings will not reflect the potential contribution of the intervention to development. This provides another key reason for not only tracking, but understanding progress and development trajectories through adaptive learning and management.

4. Positioning for development impact

E4D focuses on assessing whether the intervention is well positioned for development impact.

Under close scrutiny, the widely used criteria of relevance, effectiveness, and impact are not sufficient to foster a good understanding of how well an intervention has been, or is being positioned to enable it to contribute to development impact. In addition therefore, it is necessary to examine the intended and actual nature and scope of the intervention and of its outcomes and impacts. This will help determine the coverage and significance of the intervention in relation to a specific development context. Such analyses will focus on, for example:

- **The intervention design:** How well are the intended outcomes and impacts connected to a given model of development? Targeting immediate, intermediate or root causes of a problem? Enabling or preventing change?
- **Implementation:** Initiated at the right time, given the society, country ►►



- ▶ or region's development priorities, needs and/or strategies? Signs of (potentially) catalytic actions that might, or have been shown to accelerate progress towards development impact? Of beneficial (or otherwise) reinforcing feedback loops, or of tipping points reached?
- **Intended or actual changes:** Incremental, developmental or transformative? Dealing with large societal or whole systems change, or modest directional, incremental nudges towards a certain result? Actions are causing, or have caused impact ripples beyond expectations? The size, scope and coverage of the (intended) changes, given what is required for development (For example, is "two percent improvement" enough to justify the intervention? Is there sufficient coverage of the population and its ecosystem? Have vulnerable or marginalised groups been sufficiently reached? Are resources adequate and appropriate to enable the enable the desired changes?).

5. Managing risk towards sustained (development) impact

E4D focuses on managing the risk that desired outcomes and impacts will not be fully achieved, or might not sustain in support of development.

A critical point is that if progress, impacts and development contributions are to be sustained, we have to bring much more nuance to how we evaluate "sustainability". It is usually included as criteria in development evaluation, yet almost always only superficially addressed. It is inappropriate to judge an intervention as "successful" if the emerging or achieved positive outcomes and impacts do not sustain, or spontaneously allow other

desirable positive changes linked to development to emerge.

This requires a much stronger focus on risk management to ensure the best chance of success, based on the dual strategy of (i) strengthening the potential positive (potential or emerging) development results, while (ii) attending to those factors that might prevent such results from being achieved or sustained.

The following then become important evaluation foci:

1. **Building strength through complementarity:** In low-income countries where development indicators are weak, development efforts have to proceed systematically, coherently and in synergy. In other words, activities in each intervention (and different interventions) should be executed in a certain order, in sequence or in parallel, building upon and enhancing one another for the most impact. In order to reflect an upward development trajectory in a society, this has to take place over decades, within and across ecosystems, integrating economic, socio-cultural and environmental aspects.

Evaluators therefore have to assess whether sufficient attention has been paid to the dynamics that result from the interdependence of interventions, actions, goals, etc. (*Nilsson et al, 2016*). This means attending to the extent to which an intervention or set of interventions has made use of the power of synergistic effects that make the whole more effective than the sum of the parts.

2. **Neutralising any negative side effects.** We cannot assess a development intervention as "successful" or having achieved "impact" unless we attend to unanticipated ►►

- ▶ negative consequences, outcomes or impacts that might have significantly reduced or even nullified the positive outcomes and impacts (or that have potential to do so in future).

We have to understand whether enough has been done by the intervention designers and implementers to neutralise such negative side effects. We also have to search methodically for both positive and negative side effects, as well as for those influences that hinder or obstruct, or facilitate and enhance success, and use this information in our assessment.

3. **Delivering and retaining impact: Making it “stick”.** Implementation has to deliver development outcomes and impacts well, and in a manner that will help positive change to sustain, or foster other positive impacts. This justifies evaluative foci on effectiveness and efficiency – both well-known DAC evaluation criteria. It also justifies the increasing engagement with adaptive learning and management. This implies that we have to assess the extent to which adaptive learning and management have been properly applied in an intervention to ensure its adaptation where desirable, yet still within a system that is accountable for realistic, appropriate results.

Furthermore, in order to ensure that positive changes “stick”, we also have to attend to what we can call – for lack of a better term – the legitimacy of an intervention. This will be determined by societal and stakeholder values and perspectives, for example considering whether stakeholder voices, gender or

vulnerable groups have been sufficiently attended to in the design and implementation of the intervention, or aspects of societal culture.

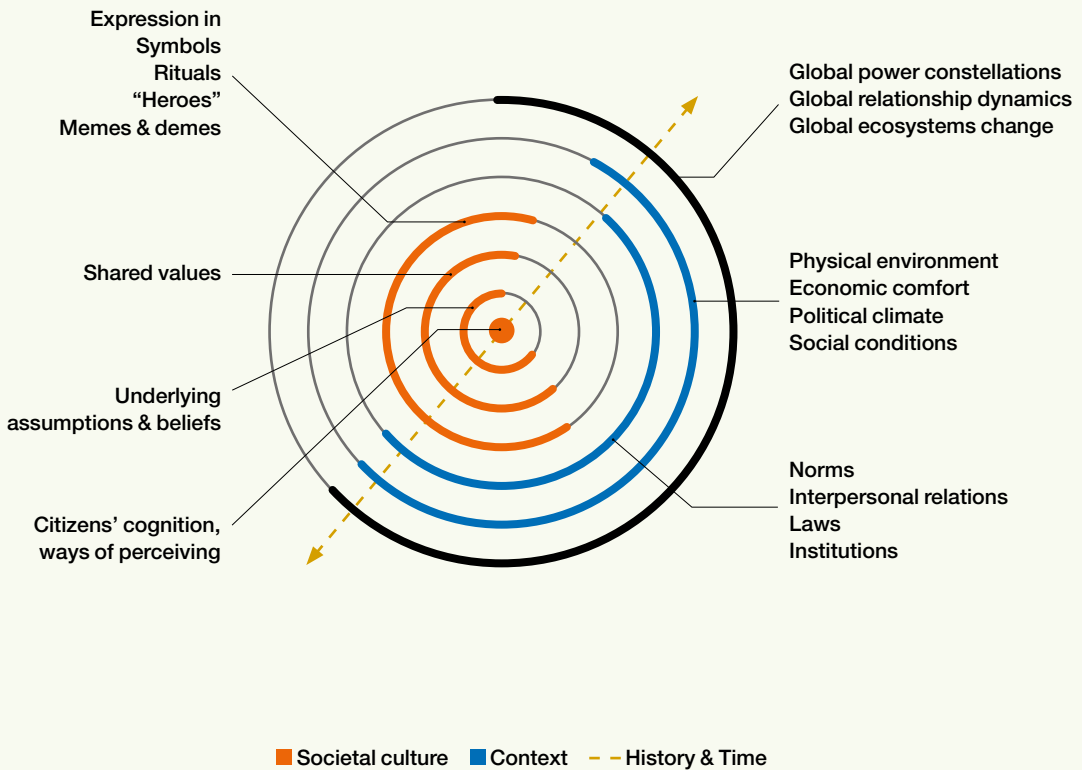
A critical aspect for consideration in E4D is the coevolution of societal culture and context (Figure 1). This shapes a society’s mental models and patterns of thinking and behaviour, i.e. its psyche, and its disposition and response towards a specific intervention (Ofir, 2016). In turn, an intervention can also affect the societal culture. This has significant implications for development, yet is seldom taken into account when we evaluate. If development impacts are to sustain, this aspect deserves to be at the centre of attention of the evaluation community.

Towards an E4D framework

There are interesting similarities between the growing interest in a more holistic, complex systems based approach to development, and the growing interest in the potential of integrative medicine to address chronic disease.¹⁰ This latter trend has sparked renewed interest in traditional oriental medicine (TOM), which is a holistic, systems-based approach to health from the East that, in turn, has been based on an integrated view of a traditional system of governance. There is growing evidence that the complicated herbal mixtures and other treatments used in TOM might be more effective and suitable for managing health and preventing or addressing chronic disease than has been previously thought (Kim *et al*, 2015).

The characteristics of a holistic “evaluation for development” approach ▶

Figure 1: Elements of the coevolution of societal culture and context that shapes the psyche of a society and its patterns of behaviour



Source: Ofir, 2016.

► quite readily map onto key aspects of TOM. An effort has been made to show the similarities in diagrams in figures 2A and 2B. While concepts or frameworks from one discipline or practice should be applied with caution to another, there are reasons not to dismiss outright the apparent synergies between the two fields of work – the one dealing with an intervention in the health status of a person; the other, with an intervention in the development status of a societal group. TOM considers the human body as a miniature version of the larger universe, in line with a fractal understanding of life. It is fully aligned with the principles of systems

biology, which has emerged as central to the study of biology and medicine. Recent developments in mind-body medicine and epigenetics have also shown that our genetic make-up and biochemical reactions are greatly influenced by complex interactions within our bodies and minds, and with our behaviour as well as social and physical surrounds.

Further work is ongoing to clarify the synergies and differences between the two systems, and to consider the implications for development and its application. ►►

Figure 2A: Components of an emerging evaluation for development (E4D) framework for interventions

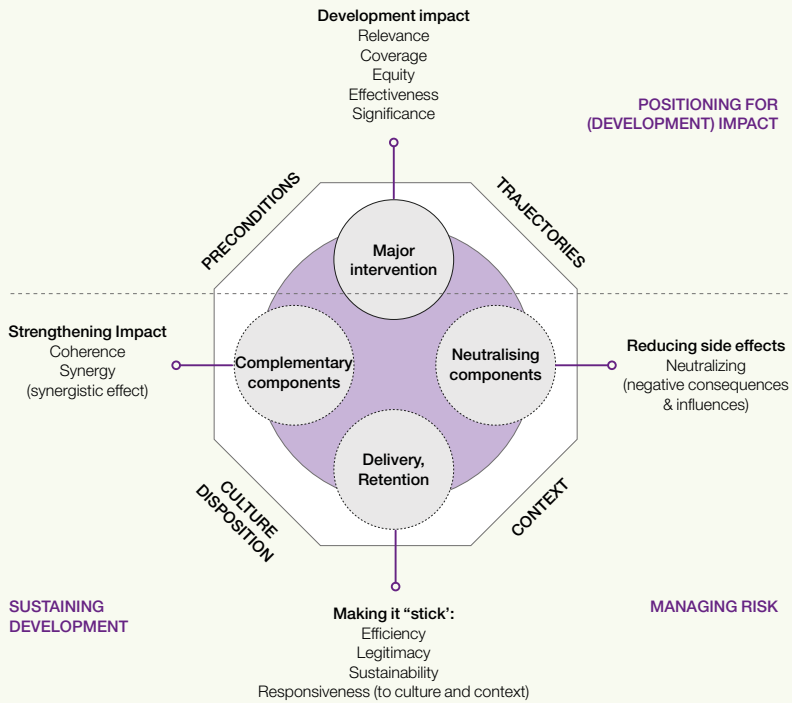
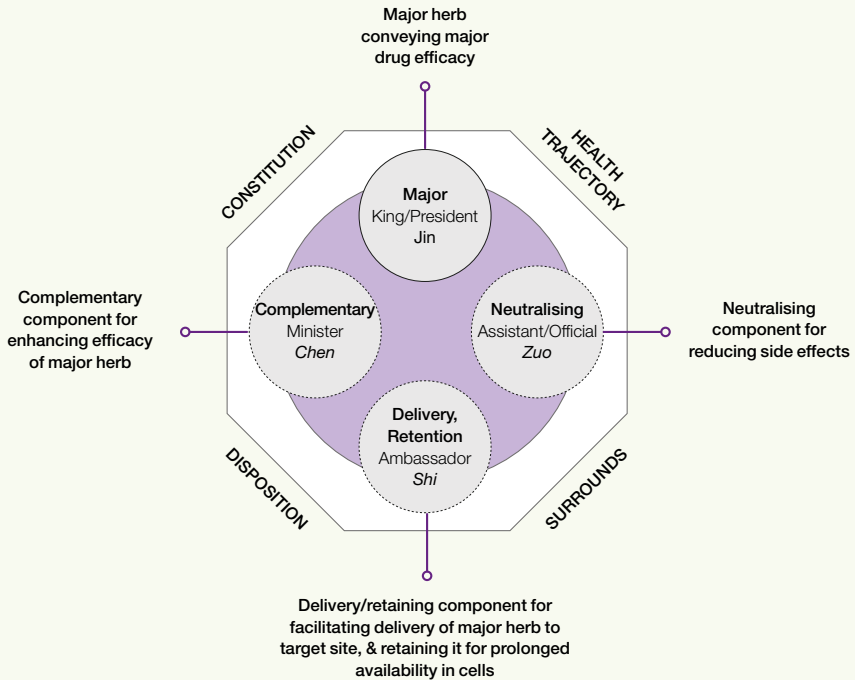


Figure 2B: Elements of the holistic systems approach of traditional oriental medicine interventions



► Revisiting our evaluation criteria

In spite of the current dynamism in the profession, evaluation has to evolve quickly to become a more powerful force in support of sustainable development. We need innovation, in particular also by those who are embedded in the Global South and are faced daily by intense development challenges.

Evaluation criteria, including the DAC criteria, have to reflect the values that we hold about what is important if evaluation is to foster development. How the criteria are conceptualised and applied can have a profound influence on our work, and on development. It is therefore important and urgent that we examine

and revise the criteria to bring renewal and nuance to what and how we evaluate.

This paper suggests one approach to do so, highlighting the implications of E4D in our search for a refined set of evaluation criteria.¹¹ Aspects for consideration are captured in figure 2a. Some of what should be done is likely going to be very hard to put into practice. But if we want to embark on Doing Evaluation Differently¹², we will require the most powerful actors in the profession – the funders and commissioners of evaluation in the Global South and Global North – to consider how best to work in concert to help ensure that our profession truly serves development. Additionally, we will require that this is fully reflected in our evaluation criteria as well as in our evaluation theories, practices and standards.

eVALU

Endnotes

- 1 Policies, strategies, portfolios, themes, programs, projects, products, events, processes, etc. – systematically planned and executed actions that result in perturbations within or across systems.
- 2 Defined in the OECD DAC Glossary as the “systematic and objective assessment of an on-going or completed development intervention, its design, implementation and results”.
- 3 It seems increasingly likely that the Fourth Industrial Revolution will reshape the nature of work, enterprises and leadership, and disrupt social, technical and ethical aspects of development. Cloud computing, smart grids, mobile web services and social media are blending with artificial intelligence, advanced robotics, and a fusion of technologies that blurs the lines between the physical, digital, and biological spheres. Power asymmetries will inevitably affect the “haves” and “have-nots”. These forces can facilitate, obstruct or destabilise development progress.
- 4 A more holistic, less selfish conception of societies’ needs and interests, and addressing humanity as a whole within planetary boundaries.
- 5 It remains essential to make the (admittedly rough) distinction between the Global North and Global South. The differences in development indicators between these two parts of the world remain too stark to ignore.
- 6 This is a preliminary list, aimed at stimulating further thought. It will be followed by a more detailed forthcoming publication.
- 7 A coherent system of biophysical and social factors that regularly interact in a resilient, sustained manner; is defined at several spatial, temporal, and organisational scales; a set of critical resources (natural, socioeconomic, and cultural) whose flow and use is regulated by a combination of ecological and social systems; and a perpetually dynamic, complex system with continuous adaptation.
- 8 A precondition is a necessary but not sufficient condition for change to happen. They include starting conditions and those that emerge as the intervention unfolds. They might include, but are not limited to, what is often called the “enabling environment”.
- 9 This is another rationale for adopting adaptive learning and management in development and evaluation.
- 10 Integrative medicine combines the reductionist approach of conventional medicine with the holistic approaches of “alternative” medicine for health management.
- 11 Potential adjustments to evaluation criteria are just briefly touched upon here, and will be expanded in a forthcoming paper.
- 12 In line with the “Doing Development Differently” initiative.

References

- Chang, Ha-Joon (2009). Hamlet without the Prince of Denmark: How development has disappeared from today's "development" discourse. Available at: <http://hajoonchang.net/wp-content/uploads/2012/01/HamletwithoutthePrinceofDenmark-revised.pdf> [Accessed 15 Dec. 2016].
- Curry-Alder, B., Kanbur R., Malone D.M., Medhora R. (eds.). (2014). International Development. Ideas, Experience & Prospects. Oxford University Press, 943.
- Heider, C. (2017). Rethinking Evaluation – Have we had enough of R/E/I/S? Blog post, January 10. Available at: <https://ieg.worldbankgroup.org/blog/rethinking-evaluation#comment-form>. [Accessed 15 Feb. 2017].
- Kim, H.U., Ryu J.Y., Lee J.O. and Lee S.Y. (2015). A systems approach to traditional oriental medicine. *Nature Biotechnology*, 33, pp. 264–268.
- Nilsson, M., Dave Griggs, D., Visbeck, M. and Ringler, C. (2016). A draft framework for understanding SDG interactions. International Council for Science. Available at: <http://www.icsu.org/publications/reports-and-reviews/working-paper-framework-for-understanding-sdg-interactions-2016/SDG-interactions-working-paper.pdf>. [Accessed 12 Sept. 2016]
- Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, 325, pp. 419–322.
- Ofir, Z. (2016). Evaluation and the nation state: where culture, context and development meet. Available at https://www.europeanevaluation.org/sites/default/files/ees_newsletter/ees-newsletter-2016-09-september-r10-web.pdf
- Picciotto, R. (2015). *The Fifth Wave*. Social impact evaluation. Published by The Rockefeller Foundation. Available at: <https://assets.rockefellerfoundation.org/app/uploads/20160628150404/The-5th-Wave-Social-Impact-Evaluation.pdf> [Accessed 10 Oct 2016].
- Ramalingam, P. (2013). *Aid on the Edge of Chaos*. Oxford University Press. 440.

Author's profile

Zenda Ofir is a South African evaluation specialist. With a PhD in Chemistry, she was Senior programme manager at a South African national science council and Director Research at University of Pretoria before turning to evaluation. Since 2000 she has worked from local to global level on evaluation assignments, focusing on Africa and Asia, and advised organisations such as GAVI, UNDP, CLEAR-AA, IFAD, AGRA and the Rockefeller Foundation. She is a former President of AfREA, former Vice-President of IOCE, and former Board Member of AEA. She serves on the editorial boards of Evaluation and Program Planning and the African Evaluation Journal. She has been a visiting professor at the University of Hiroshima, has taught at the United Nations University in Tokyo, and has been designated Honorary Professor at Stellenbosch University, South Africa, an appointment that recognises professional eminence in a career outside the university.

