

Have Development Evaluators Been Fighting the Last War... And If So, What is to be Done?

Robert Picciotto

Abstract What are the right evaluation methods and what evaluation strategy adapted to contemporary development issues should such methods serve? This article argues that a range of methods and tools must be drawn upon to achieve high-quality evaluation standards and that even the best methods cannot compensate for a misguided strategy. For both sets of reasons, it is time for the current craze for experimental evaluation methods in international development to give way. In its place a mix of evaluation methods informed by a relevant evaluation research agenda should be adopted and put at the service of evaluation policy directions that reflect contemporary development prospects and dilemmas.

Evaluation cannot hope for perfect objectivity but neither does this mean that it should slump into rampant subjectivity.

Ray Pawson, *The Science of Evaluation: A Realist Manifesto* (2013)

1 Introduction

A research agenda is relevant only if guided by reasoned answers to two evaluative questions. First, the summative question: is the agenda soundly grounded on an objective examination of the state of the art? Second, the formative question: does it address emerging and legitimate policy concerns? Accordingly, this article is in two parts. Following a background section that sets the stage, the first half of the article addresses the potential and limitations of experimental methods that have come to dominate the development evaluation domain. The second half looks ahead: it draws the implications of a rapidly evolving international development context for the impact evaluation research agenda.

2 Setting the stage

We are all creatures of habit. For individuals (and even more so for organisations), entrenched ways of acting are frequent. Yet management experts have long stressed the risks associated with repetitive behaviour and fixed operating

protocols, especially if the operating environment is uncertain or unstable. As the management guru Peter Drucker famously opined, the biggest curse for any business is 20 years of success since it induces complacency and arrogance (Ashkenas 2012).

Development evaluation is tasked to examine complex problems in diverse and ever-changing operating environments. In such contexts, sticking to a single evaluation model and rigidly complying with the strictures of a single method (however fashionable) is a threat to evaluation quality. Similarly, evaluation programming that does not take account of emergent risks and evolving priorities wastes scarce evaluation resources. Equally, the reiteration of old, tired and trite 'lessons' attracts derision.

Unfortunately, evaluators are peculiarly vulnerable to path dependent ways of thinking and working. To be sure, recollecting in tranquillity about past performance constitutes a healthy reality test for policymakers all too often wedded to the politics of good intentions rather than the discipline of accountability, transparency and results. But retrospective evaluation is not the only contributor to sound decision-making.

2.1 Impact definitions matter

Assessing impact as defined by the 3ie organisation (attribution of effects to an intervention) (White 2009) has major implications for the social utility of evaluations. A focus on attribution concentrates on the 'what' rather than the 'why' and the 'how'. It also evades the question of who was responsible for the success or failure of an intervention.

Thus as a summative process it is not always supportive of accountability. Nor are its formative benefits always self-evident. A programme that did not 'work' in the past may be fine-tuned or transformed so that it 'works' in the future. Indeed one may have been wrong by being right too early; what worked well yesterday may not work tomorrow and what did not work yesterday may conceivably (and with some luck) work tomorrow.

2.2 The past is not always prologue

Indeed, opinionated evaluators exclusively focused on the past may end up doing more harm than good. Extrapolation of past lessons into the future is a tricky business. Highly successful approaches that have stood the test of time can suddenly lose their effectiveness. Just as internal validity does not necessarily translate into external validity across country contexts, the lessons of the past do not necessarily hold for the future. History rarely repeats itself even though it often rhymes.

This explains why experienced evaluators exercise great care when asked to produce recommendations. They know that information asymmetries regarding the force field within which decision-makers operate do not favour external observers. They appreciate that operational objectives and implementation processes need to reflect expected changes in the operating and authorising environment. Predicting the future is a risky endeavour.

Waiting to fix something until it is broken is not always a good management practice, especially in a competitive environment. Fighting the war with old tactics and archaic weapons is doomed especially when the terrain is new, adversaries have changed and/or technologies have evolved. Conversely, the mistaken decisions of the past may turn out to be inspired in a transformed context: one way of being wrong is being right too early.

And so it is in development... and in evaluation. The risks of path dependence affect the

evaluation discipline itself – as well as the development enterprise that evaluators are called upon to guide. Circumstances change. Coming to terms with the fact that development takes place in a changing, complex and uncertain world has fundamental consequences for evaluation concepts and methods. This is fundamentally why 'fighting the last war' (the war of ideas about how to assess development effectiveness) is a clear and present danger in development evaluation.

2.3 It is time to move beyond the paradigm wars

Looking back, evaluating the impact of discrete aid interventions has dominated recent campaigns aimed at capturing the commanding heights of development evaluation. In this field of evaluation practice skirmishes between qualitative and quantitative evaluators are still erupting. But attribution of effects to individual interventions should be only one of the questions to be tackled at this particular juncture in development history. In truth, diminishing returns, argumentative gridlocks and overall fatigue have set in on this particular battlefield. It is time to move on and open new fronts towards victory for our fledgling discipline (Picciotto 2012).

To be sure, the intellectual contest about causal inference has long been and will continue to be a focus of methodological inquiry in social science and evaluation research. But we should finally acknowledge that evaluators, social scientists and philosophers of science have thoroughly explored its various dimensions. The topic has been at the centre of the evaluation research agenda since the origins of the evaluation craft when the development idea took shape and evaluation pioneers forged the basic tools of the discipline under the aegis of Donald Campbell's experimenting society.

The long-standing and highly charged debate of this contested terrain was recently revisited in a recent landmark study funded by the Department for International Development (Stern *et al.* 2012). It is a balanced and definitive state of the art review. It consolidates the fragile truce that has prevailed ever since the Network of Networks on Impact Evaluation delivered its initial verdict in an authoritative report dated 2009 (Leeuw and Vaessen 2009). By now, the potentials and limitations of the experimental approach have been thoroughly probed and the hidden

assumptions that underlie experimentalism have been fully unearthed and finely dissected along the lines of what follows.

3 The pros and cons of experimental methods

In the right circumstances and with the right skills, experimental methods can help establish causality by providing a plausible measure of what results would have been observed had the intervention not taken place. But how prevalent are such situations in the real world of international development?

3.1 The potential of randomisation

Randomised control trials aim to achieve comparability between control and treatment groups as a result of random selection of beneficiaries and non-beneficiaries drawn from the same population through an explicit chance-based process. Unbiased allocation means that the probability of ending up in the control group or the treatment group is identical. This approach is ideal for tackling *selection bias*.

Such bias prevails when comparing impacts on substantially different sets of beneficiaries. In naive evaluations it induces false attribution of observed results by ignoring vital differences in the known or unknown characteristics of the treatment and non-treatment groups. The problem arises for example – and is often the case – when those who access the programme are richer, more powerful, more motivated or more educated.

Random assignment to the treatment and non-treatment groups from the same population group ensures that, except for chance fluctuations, the impact of the intervention can be reliably ascertained by comparing outcomes among the two groups by ensuring that all the other factors that may affect outcomes are identical except for stochastic errors.

To ascertain the reliability of the finding, statistical testing techniques are available to determine the range of confidence that one may safely attribute to the result (i.e. the role that pure chance associated with the randomisation process may have played).

Thus, randomised control trials enjoy a singular advantage: in the right context and with the right skills and large populations they allow

evaluators to establish a measure of statistical significance to evaluation findings.

3.2 The limits of experiments in society

What then, given these formidable advantages, is the applicability of randomised control trials for assessing the impact of development interventions? Undoubtedly, experimental methods are an integral part of the evaluator's toolkit. But we should also acknowledge that they do not work well unless the operating environment is relatively stable and the intervention being evaluated is fixed and well defined.

The melancholy fact is that these conditions are not often met in international development. Experiments are redundant when no other plausible explanation for the results observed is available. They may not constitute a feasible option. For example, it is not possible to randomise the location of infrastructure projects (Ravallion 2009). Nor can experimental methods be used when no untreated target group can be identified, for example when the intervention aims at full coverage in a region or a country.

Experiments may not even be decisive in establishing attribution. Poverty is not a disease and development aid is not a drug. Society is not a scientific laboratory. Reflexivity is a common feature of social phenomena. Strong feedback is apt to induce instability. Only if the treatment group and the control group and the process that affects each are strictly identical (except in terms of cause and effect) can inferences be established with confidence. Internal validity may also be jeopardised by latent and unobserved causal factors that are not taken into account when constructing the treatment and control groups.

Nor is external validity the *forte* of experimental methods. Even where experiments are appropriate they may not meet the needs of policymakers who may be vitally concerned not so much with what worked in a trial experiment, but with whether they are likely to keep working when replicated or up-scaled in a diverse and volatile implementation environment (Cartwright and Munro 2010). Programme size, structure and context matter a great deal in shaping the outcome of development activities.

Depriving members of the control group of a useful treatment based on a selection process

perceived as capricious and arbitrary can be considered discriminatory and may even be illegal. In some jurisdictions, no comparison group is allowed to receive any treatment that is less than the best currently available. Nor is it usually considered ethical to induce members of a treatment group to participate in an intervention that may have negative side-effects. Informed consent procedures and other ways to minimise ethical lapses often introduce the very selection bias that the method was supposed to guard against.

Finally, even where experiments make sense for assessing attribution, they require superior skills, large studies, large samples and specialised quality assurance arrangements. These pre-requisites may not be available in real life situations and they may not translate into an economic use of scarce evaluation resources. They may inhibit resort to cheaper and more effective evaluations. They may also hinder fulsome participation of aid recipients in the evaluation process by shifting the control of sophisticated impact evaluation to well-endowed universities and thinktanks located in developed countries.

3.3 *The promise of pluralistic methods*

The flip side of the above limitations is that, without a theory that has survived a range of validity tests there is no credible causal explanation that adds to development knowledge. A deep understanding of how a particular programme operates is critical and the validity of the theory on which it is predicated must be established.

Securing an adequate understanding of causal relationships and identifying the rival explanations that need refutation call for substantive knowledge of the intervention, its design, its implementation protocols and the incentives of programme participants and beneficiaries. This is the unchallenged province of qualitative methods.

Qualitative methods are not a panacea and they need to be complemented by quantitative methods. But they can be rigorously applied. They have the singular merit of examining who did what; how things actually happened and why. They involve participation, observation, analysis of text-based information, village meetings, open ended interviews, etc. Of course, qualitative data

collection requires careful coding and systematic quantification in order to be analysed.

Quality can and must be quantified. It is also important to figure out why things did *not* happen. Thus Rob Van den Berg and Christine Woerlen have advocated theories of ‘no change’ that address the question of why programmes that should have worked did not work due to implementation problems or exogenous shocks (Van den Berg and Woerlen 2013).

3.4 *Alternative approaches*

In conclusion, while large quantitative studies are invaluable, rich qualitative descriptions of individual cases should have an equally prominent place in the evaluator’s toolkit. Whereas experimental methods are shaped by data, qualitative, theory-based approaches are shaped by questions of special interest to stakeholders and by the assumptions embedded in the design of programme and project intervention (Bamberger, Rao and Woolcock 2010).

Rigorously conducted case studies are exceptionally well equipped to illuminate the reasons for success or failure of achieving intended effects (as well as the extent and nature of unintended effects). In particular, they can help to distinguish between design problems or conflicting stakeholders’ values or run-of-the-mill implementation weaknesses related to resource constraints, skills shortages or misguided operating protocols.

This is why many experienced and successful evaluators have gone through their whole career without ever using a randomised control trial. More often than not, the evaluation questions of interest to stakeholders require a different approach than randomisation. For example, contribution analysis may be better adapted to addressing issues of accountability for interventions implemented through partnerships. Other approaches (process tracing, qualitative comparative analysis, social simulation techniques) are also being explored.

More generally, evaluation commissioners may be interested less in whether interventions work than in *why* they do, *whether* design or implementation problems explain observed setbacks or *who* among development partners is responsible for particular outcome characteristics. To be sure,

experimental methods have many powerful statistical features that other evaluation designs cannot easily match.

On the other hand, a threat to good evaluation management is overinvestment in a single technique.

A wide variety of tools exists to simulate a counterfactual short of randomisation. Regression and factor analysis, quasi-experimental designs, multivariate statistical modelling, participatory approaches, qualitative impact assessment, beneficiary surveys and sampling, general elimination methodology, expert panels and benchmarking are often used where experiments are not feasible or cost-effective.

To be sure, all of these methods have their own limitations but triangulating among them offers scope for greater evaluation rigour. More often than not, a tool fulfils only the function or functions that it was designed for. Privileging public interventions that are evaluable through experimental methods encourages the selection of simplistic programmes and projects that may not be fit for purpose and/or promote avoidance of critical evaluation questions. Mixed methods tailor-made to individual cases is the key to evaluation quality. This is why disruptive skirmishes in a paradigm war that has long been settled should be avoided.

The jury is in: randomisation is mostly suited to simple interventions with easily identified participants and non-participants and where spillover effects are not likely to bias the results. It is poorly suited to the evaluation of adaptable, complicated or complex programmes in unstable environments. Yet, this is what development evaluation is mostly about. It is also where knowledge gaps are the deepest.

3.5 Complexity and systems thinking

All evaluation tools, including randomised control trials, are just tools. They should not be allowed to dominate what is first and foremost a creative, analytical and participatory process. From this vantage point the new perspectives offered by complexity science and systems thinking are promising as well as challenging.

This family of approaches embraces pluralism and inclusivity in methods. The various models

embedded in systems thinking are still in their infancy. But they hold considerable attraction as highlighted by other articles in this issue of the *IDS Bulletin*: they adopt networks as the privileged unit of account. They explore inter-relationships and engage with multiple actors. They use concepts, symbols and diagrams to guide evaluative inquiry. They probe the multiple loops that illuminate a changed preoccupation from 'attribution' to 'contribution'.

Finally, they call on all evaluation traditions – including theories of change, realist evaluation and critical evaluation models focused on the overt and covert pressures that shape society. Looking ahead, both complexity theory and systems thinking working in tandem may well induce a convergence between development evaluation and developmental evaluation. But we are still at an early stage in a long and arduous intellectual journey. We will have to endure a long twilight intellectual struggle to get to the top of the hill identified by complexity and systems thinkers.

4 Towards a new development evaluation research agenda

Laying the foundations of a research and practice agenda related to impact evaluation evokes a deliberate focus on results and methods. Continuing with the warfare analogy, methods have to do with tactics. They should be judiciously selected and skilfully deployed. But they should also serve a strategy that fits the needs of the times.

Having come to terms with the reality that most high-level policies, programmes and projects that are now privileged by international development agencies are not evaluable through randomised treatment, the new development evaluation research agenda will need to explore new evaluation frontiers. But methods only work if they are selected to serve the right strategy, i.e. a strategy that makes sense in the current operating and authorising environment.

A relevant development evaluation strategy for our troubled times should meet three criteria. First, it should put *values* at the very centre of the strategy. Second, it should accurately reflect the *new development agenda* likely to emerge beyond the 2015 finishing line of the Millennium Development Goals (MDGs). Third, it should

address the realities of an aid enterprise in turmoil and a reconsideration of the ultimate goals of development cooperation.

4.1 Values should come first

In rich and poor countries alike, the problems of others have become our own. From the clamour of the Arab Spring to the labour unrest of South African mining communities and the Occupy Wall Street protests that have flared around the world, the social arrangements and governance practices that allow a few individuals to accumulate enormous wealth while depriving large segments of the population of basic necessities have spawned widespread popular anger and resentment. Unless evaluators respond to the popular clamour for more equitable and sustainable policies geared to social inclusion and solidarity, the evaluation discipline will gradually lose its legitimacy.

The 2008 global economic crisis has brought into focus the unfair social arrangements and dysfunctional governance practices that have allowed a few individuals to accumulate enormous wealth while depriving large segments of the world population of decent livelihoods. The economic, political and social mechanisms that underlie growing inequality have recently come under close scrutiny (Piketty 2014).

Differences in income and wealth attributable to effort, skill or entrepreneurship are not widely resented. But public indignation and anger spread and social cohesion is undermined when distorted rules of the game, predatory economic behaviour or unethical business practices are richly rewarded. The richest 0.5 per cent hold well over a third of the world's wealth while 68 per cent share only 4 per cent (Credit Suisse Research Institute 2010). According to Joseph Stiglitz (2012: xi): 'there are moments in history when people all over the world seem to rise up to say that something is wrong'. We are living such a moment.

4.2 Evaluation is not value neutral

This is why evaluators will have to give greater weight to ethical values and social justice imperatives. Over the past two decades goal-oriented, client-controlled evaluations have contributed to the timidity of evaluation agendas. In these extraordinary times, anchoring evaluation in commitments to democracy, social

justice, empowerment and equality has become imperative. Delivering evaluation results matters of course, but whose results and to which actor(s) should they be attributed?

Methodological biases have given pride of place to linear methods that do not contribute meaningfully to the assessment of the full coverage policies that hold the key to social equity and public welfare. Development evaluators have been mostly preoccupied with the use of aid funds by developing countries while aid delivery mechanisms and non-aid policies shaped by donor countries have escaped similar scrutiny. More attention has been devoted to assessing developing countries' performance than to the fulfilment of donor countries' reciprocal partnership obligations.

Given that most evaluators are contractually dependent, many of them have been prone to frame their evaluations to meet programme managers' needs and concerns rather than those of citizens. They have found ample justification for their supine stance in the business management literature and utilisation-focused evaluation textbooks. Yet evading or downplaying the social and summative dimension of evaluation in order to make evaluation findings palatable does not serve to make authority responsible or responsive to the public interest.

The American Evaluation Association Guiding Principles¹ are extraordinarily influential. They urge evaluators to meet legitimate clients' needs whenever it is feasible and appropriate to do so.² The proviso that follows this principle notwithstanding,³ the Principles do not address the ethical responsibilities of evaluation commissioners.⁴ This helps to enshrine their primacy in evaluation governance. The resulting client emphasis and the comparative neglect of political and evaluation governance issues have contributed to the dominance of a market-based evaluation system that is not always compatible with the public interest.

Serving programme managers and decision-makers is what management consultants do. Evaluators have a broader remit: they must also serve the public interest. Giving pride of place to the utilisation of evaluation results is socially useful only where decision-makers have not been wittingly or unwittingly captured by vested

interests. Conversely, a utilisation-based evaluation doctrine that gives little weight to inequities can be socially harmful and undermine evaluation independence.

Bringing values back from the cold implies an ethical evaluation stance somewhat at variance with the client-controlled, goal-based approaches that currently dominate the field. By contrast, development evaluators should seek inspiration in contemporary ideas about social justice. They should reach out to their social sciences colleagues in order to benefit from recent and far-reaching policy research findings about inequality. They should develop metrics consistent with the pursuit of more equitable, inclusive and environmentally sustainable growth. Last but not least, they should give higher priority to the global dimensions of policies and programmes.

4.3 Adapting to an emerging development context

Without the right values evaluation is not likely to have merit or worth. But irrelevance also lurks if development evaluators do not take account of the universal and legitimate aspirations embedded in the emerging development consensus. Evaluation strategies and programmes should reflect recent shifts in policy paradigms. By specifying a wide range of socioeconomic indicators, the MDGs displaced economic growth as the dominant objective of development. The Sustainable Development goals will put even more emphasis on socially inclusive and environmentally sound development strategies.

This shift in emphasis is consistent with the recognition that national income growth is not a good proxy for economic and social performance. Gross national income and traditional growth measures fail to capture highly valuable services provided within the household. Nor do they take account of environmental losses, prevailing social inequities or the human insecurities associated with unbridled growth.

Evaluators should also keep their sights on the chronic challenge of global poverty reduction. More than 70 million fewer people than estimated in 2008 when the crisis began are projected to escape extreme poverty by 2020 (World Bank 2010a). Poverty will not be made history any time soon unless the poverty threshold is defined at an absurdly low level.⁵

Food security is back on the international agenda: with about 868 million people still going to bed hungry every night (FAO 2014). Health problems are not letting up: developing countries still account for 93 per cent of the worldwide burden of disease and only 11 per cent of global health spending (Schieber and Maeda 1999).

Current energy and natural resources intensive policies are not sustainable. Five planets would be needed to accommodate all countries at current US standards of living. The spectre of global warming is haunting the poorest and most vulnerable countries. The world's 1,210 billionaires have a combined fortune of US\$4.5 trillion – over half of the net worth of three billion adults (*Forbes Magazine* 2011). These glaring inequalities combined with the unexpected collapse of the loosely regulated global financial system have thoroughly shaken policy complacency among voters, policymakers, as well as academics.

By now, the dominance of the Western economic model is being reconsidered: adjustment problems similar to those endured by debt-burdened developing countries in the mid-1980s are now plaguing Western countries. The North–South model that characterised international relations in the second half of the twentieth century has become anachronistic. China's gross domestic product (GDP) already exceeds Japan's, while Brazil's will overtake that of France and the United Kingdom by the middle of this decade.

4.4 Towards new metrics

Looking ahead, it is quality growth (rather than growth per se) that will constitute the overarching economic, political and ethical imperative of the contemporary development enterprise. But how to achieve it is not self-evident in the wake of an unprecedented financial crisis that has turned decades of economic orthodoxy on its head. The goals, principles and practices that have long governed development cooperation cannot be a good guide for the future. Thus, development thinking is still evolving.

The reality of poverty encompasses economic, social, cultural, institutional and subjective factors that reach well beyond material deprivation. Indeed, Amartya Sen, the Nobel

laureate, has equated development with freedom (Sen 1999). Conceptions of human development should at long last be informed by what poor people have to say about their own distinctive predicaments. At the same time, the profile of risk management should rise given the volatility and the insecurities of the interconnected global economic system. Evaluation should adapt to the emerging development paradigm.

Beyond income, health conditions, housing, employment, social cohesion, environmental quality and quality of life, indicators will have to be identified through collaborative research. Taking such dimensions of human progress into account will provide more reliable assessments of policy and programme performance within and across countries. Of course, the time and resources needed to secure broad-based agreement on a comprehensive set of human wellbeing indicators (let alone measure them and use them for policymaking) is a long-term endeavour involving statistical agencies, international organisations and academic institutions.

As a synthesis of converging ideas regarding global development goals beyond 2015 the three dimensional (3D) model crafted in IDS by Allister McGregor and Andy Sumner (see McGregor and Sumner 2010) offers a timely analytical tool that captures the material, relational and perceptual characteristics of human aspirations and social progress and provides a convenient framework for assessing development interventions, whether focused on improved capabilities or more favourable enabling conditions.

For all agencies concerned with development today the matrix below points to the diverse and complementary evaluation disciplines, approaches and concepts that will have to be marshalled to do justice to the holistic conception of development embedded in human wellbeing aspirations. Here again, complexity and systems thinking may help solve the riddle.

Finally, development evaluation should take on board the fact that the aid architecture is becoming more and more fragmented and that the consensus that was laboriously constructed at the turn of the century to generate a momentum in support of the MDGs is unravelling.

4.5 Taking account of a new aid architecture

It is an open secret that development aid was badly misused in the Cold War confrontation with the Soviet Union and that it is once again conceived primarily as a soft power instrument at the service of donor countries' diplomacy. To be sure poverty reduction is still vigorously promoted by principled civil society actors but since the turn of the century and given a political climate dominated by social insecurity and budget austerity, official development aid has become more subservient to the national security and commercial goals of the foreign policy establishment.

Aid is being used systematically to prop up allies, open new markets or secure access to natural resources. At its worst, aid has been reduced to one of the weapons deployed in a global war on terror that has no end and fiercely resists objective assessment. Given these mixed goals and quite apart from the tectonic shifts that have shaken the aid industry since the 2008 financial crisis, it is less and less credible to evaluate aid programmes against the admirable intentions that politicians are prone to articulate when they seek to explain to their domestic constituencies why it makes sense to use scarce budgets for overseas interventions when domestic social programmes are being cut.

Finally, aid interventions today are increasingly channelled through large, diffuse and adaptable country-based programmes or through vertical multi-country initiatives involving multiple actors. The broad-based aid coalitions that are increasingly operating across borders to tackle 'problems without passports' must contend with the rules of the game set by multi-sector networks. Complex political equations underlie their operations. They cannot be solved through abstract and studiously neutral experiments.

4.6 Shifting gears from aid evaluation to development evaluation

The Busan Declaration promotes a shift in the development paradigm from aid effectiveness to development effectiveness. This is nothing short of revolutionary for the evaluation agenda. Gone are the days when aid interventions could be evaluated as discrete, isolated and evaluable through reductionist approaches. Development cooperation is now conceived as a web of partnerships involving state-based structures,

Table 1 Human wellbeing and evaluation

Evaluation characteristics	Material wellbeing	Relational wellbeing	Perceptual wellbeing
Major discipline	Economics	Sociology	Psychology
Dominant evaluation approach	Cost benefit analysis	Participatory evaluation	Empowerment evaluation
Investment focus	Physical capital	Social capital	Human capital
Main unit of account	Countries	Communities	Individuals
Main types of indicators	Socioeconomic	Resilience	Quality of life

Source Adapted from McGregor and Sumner (2010).

private corporations and civil society groups. Systems thinking appears better suited to figuring them out than simple linear mental models.

Underlying the new development order is the growing interconnectedness of nations and societies and its corollary: development cooperation beyond aid. Non-aid links have become major mechanisms of resource transfer that are dwarfing the ‘money’ impact of aid and creating new and powerful connections between rich and poor countries (as well as among poor countries):

- Developing countries’ exports (about US\$5.8 trillion) are 45 times the level of official aid flows.⁶
- Remittances from migrants (US\$283 billion) are 2.2 times as large as aid flows.⁷
- Foreign direct investment (US\$594 billion) is 4.6 times as large as aid flows.⁸
- Royalty and licence fees paid by developing countries to developed countries (US\$27 billion) are over a quarter of aid flows.⁹
- The huge damage to developing countries caused by climate change as a result of OECD countries’ unsustainable environmental practices is getting worse given rapid growth in emerging market countries (UNFCCC 2007).

These powerful transmission belts of globalisation mean that the footprint of development cooperation is far more significant than the aid footprint. From an evaluation perspective the full agenda of policy coherence for development is

now open for evaluative scrutiny. An early start is the Commitment to Development index issued by the Centre for Global Development (Barder and Krylová 2013). It demonstrates that aid quality and quantity have only a limited effect on the overall country rankings.

Consequently, evaluation at the higher plane of global policy has much to contribute to the redesign of a development cooperation agenda that reaches beyond aid. In the process of evaluating interventions from a ‘policy coherence for development’ perspective,¹⁰ evaluation methods will have to be informed by a wide range of disciplines ranging from the political sciences, to the sociology of regulatory networks, the economics of institutions and the organisational management literature.

5 Conclusions

Looking ahead, a relevant research and evaluation agenda should emphasise ethical standards and democratic values. It should reflect the human wellbeing paradigm that is gaining currency in development circles. It should come to terms with the realities of the fractured aid architecture and the imperatives of a development cooperation agenda that reaches well beyond aid. These policy directions imply greater independence in evaluation processes, more attention to the political and social aspects of development policies and programmes and full resort to mixed and inclusive methods illuminated by the emerging insights of complexity science and systems thinking.

Notes

- 1 For a full description, please see: www.eval.org/p/cm/ld/fid=51.
- 2 Principle E 4: 'Evaluators necessarily have a special relationship with the client who funds or requests the evaluation. By virtue of that relationship, evaluators must strive to meet legitimate client needs whenever it is feasible and appropriate to do so.'
- 3 Principle E 5: 'Evaluators have obligations that encompass the public interest and good. These obligations are especially important when evaluators are supported by publicly-generated funds; but clear threats to the public good should never be ignored in any evaluation.'
- 4 This contrasts with the ethical guidelines of the United Kingdom Evaluation Society, which distinguish between the accountability of evaluators, commissioners, participants and those involved in self-evaluation in organisations (www.evaluation.org.uk/about-us/publications).
- 5 The number of people living below the (miserly) international poverty line of US\$1.25 per day fell from 1.82 billion to

References

- Ashkenas, Rob (2012) 'Kill Your Business Model Before It Kills You', *Harvard Business Review (HBR) Blog*, <http://blogs.hbr.org/2012/10/kill-your-business-model-befor/> (accessed 3 August 2014)
- Bamberger, Michael; Rao, Vijayendra and Woolcock, Michael (2010) *Using Mixed Methods in Monitoring and Evaluation, Experiences from International Development*, Policy Research Working Paper 5245, Washington DC: World Bank Research Group
- Barder, Owen and Krylová, Petra (2013) *Commitment to Development Index 2013*, Center for Global Development Brief, www.cgdev.org/publication/commitment-development-index-2013 (accessed 3 August 2014)
- Cartwright, Nancy and Munro, Eileen (2010) 'The Limitations of Randomized Controlled Trials in Predicting Effectiveness', *Journal of Evaluation in Clinical Practice* 16.2: 260–66
- Credit Suisse Research Institute (2010) *Global Wealth Report*, Zurich
- FAO (2014) *Hunger Portal*, Food and Agriculture Organization, www.fao.org/hunger/en/ (accessed 3 August 2014)
- Forbes Magazine* (2011) 'The World's Billionaires', 9 March, www.forbes.com/lists/2011/10/

- 1.37 billion between 1990 and 2005. See Picciotto (2014).
- 6 This is the 2008 level according to the World Trade Organization (WTO). It dipped by 8 per cent in 2009 but more than fully recovered in 2010 according to the International Monetary Fund (IMF). See www.imf.org/external/pubs/ft/weo/2011/01/index.htm and www.wto.org/english/thewto_e/coher_e/mdg_e/development_e.htm.
 - 7 Remittance flows to developing countries stood at US\$283 billion in 2008 according to the World Bank (2008), which projected them to dip slightly in 2009 and more than fully recover in 2010.
 - 8 2008 estimate (World Bank 2010b).
 - 9 This is a World Bank estimate for 2007 included in the *2009 World Bank Indicators* which is compared to aid flows of \$104 billion for that year according to DAC statistics.
 - 10 Organisation for Economic Cooperation and Development (OECD) *Policy Coherence for Development: PCD in a Post-2015 World*, www.oecd.org/development/pcd/policycoherenceinapost-2015world.htm.

[billionaires_2011.html](http://www.oecd.org/development/pcd/policycoherenceinapost-2015world.htm) (accessed 3 August 2014)

- Leeuw, Frans and Vaessen, Jos (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*, Washington DC: Network of Networks on Impact Evaluation
- McGregor, Allister and Sumner, Andy (2010) 'Beyond Business as Usual: What Might 3-D Wellbeing Contribute to MDG Momentum?', *IDS Bulletin* 41.1, www.ids.ac.uk/files/dmfile/Beyond_Business_as_Usual_What_Might_3D_Human_Wellbeing_Contribute_to_MDG_Momentum.pdf (accessed 3 August 2014)
- Pawson, Ray (2013) *The Science of Evaluation: A Realist Manifesto*, London: Sage Publications
- Picciotto, Robert (2014) 'Poverty Eradication by 2030? Old Wine in a New Bottle', *King's College London, Dialogue* 07, <http://issuu.com/kclpolitics/docs/online07?e=5512974/7166752> (accessed 3 August 2014)
- Picciotto, Robert (2012) 'Experimentalism and Development Evaluation: Will the Bubble Burst?', *Evaluation* 18.2: 213–29
- Piketty, Thomas (2014) *Capital in the 21st Century*, Cambridge MA: Harvard University Press
- Ravallion, Martin (2009) 'Should the Randomistas Rule?', *Economists' Voice*, Berkeley CA: Berkeley Electronic Press

- Schieber, G. and Maeda, A. (1999) 'A Health Care Financing and Delivery in Developing Countries', *Health Affairs* 18.3: 193–205
- Sen, Amartya (1999) *Development as Freedom*, Oxford: Oxford University Press
- Stern, Elliot; Stame, Nicoletta; Mayne, John; Forss, Kim; Davies, Rick and Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluation*, DFID Working Paper 38, London: Department for International Development
- Stiglitz, Joseph E. (2012) *The Price of Inequality: How Today's Divided Society Endangers our Future*, New York NY: W.W. Norton
- UNFCCC (2007) *Impacts, Vulnerabilities and Adaptation in Developing Countries*, Bonn: United Nations Framework Convention on Climate Change Secretariat
- Van den Berg, Rob D. and Woerlen, Christine (2013) *The Importance of Negative Evidence in Evaluation and The Theory of No Change*, EO Paper Evaluation Connections, Independent Evaluation Office, Global Environment Facility, Washington DC: GEF
- White, Howard (2009) *Some Reflections on Current Debates in Impact Evaluation*, International Initiative for Impact Evaluation, Working Paper 1, www.3ieimpact.org/en/evaluation/working-papers/working-paper-1/ (accessed 4 August 2014)
- World Bank (2010a) *Global Monitoring Report*, Washington DC: World Bank
- World Bank (2010b) *Global Development Finance 2010: External Debt of Developing Countries*, Washington DC: World Bank
- World Bank (2008) *Migration and Development Brief 8*, Washington DC: World Bank