

## Chapter 3

# Mainstreaming Impact Evidence in Climate Change and Sustainable Development

**Rob D. van den Berg**

**Abstract** This chapter examines the demand for impact evidence and concludes that this demand goes beyond the experimental evidence that is produced during the lifetime of an intervention through “impact evaluations” as currently the term is used by many in the evidence movement. The demand for evidence of longer term impact at higher levels requires inspiration from an older tradition of impact evaluation and rethinking how the full range of impact evidence can be uncovered in evaluations. This is especially relevant for sustainable development which calls for a balanced approach on societal, economic and environmental issues. Climate change is a good example of this and a theory of change approach serves to identify key questions over time, space and scale to ensure that impact evidence can be found and reported throughout the lifetime of projects, programmes and policies and beyond in ex post impact assessments. Such an approach leads to mainstreaming of impact questions and related evaluation approaches throughout project and policy cycles. This chapter will demonstrate that evidence can be gathered throughout the lifetime of a project and beyond, in different geographic locations from very local to global, at different levels from relatively simple one dimensional interventions to multi-actor complex systems, up to global scales. It will thus argue for mainstreaming impact considerations throughout interventions, programmes and policies and for evaluations to gather evidence where it is available, rather than to focus the search for impact and its measurement on one or two causal mechanisms that are chosen for verification through experimentation.

**Keywords** Impact • Evaluation • Sustainable development • Climate change • Evidence

---

R.D. van den Berg (✉)

King’s College London and International Development Evaluation Association (IDEAS),  
Leidschendam, The Netherlands  
e-mail: [rdwinterberg@gmail.com](mailto:rdwinterberg@gmail.com)

### 3.1 Re-instating an Older Impact Tradition?

The debate on what constitutes impact in evaluation continues, with many in the evidence based movement<sup>1</sup> focusing on “rigorous” experiments to measure and identify what works and what doesn’t, versus participatory and democratic approaches enabling beneficiaries to state what would be relevant for them. It is important to note that both approaches, and many others, tend to focus on the here and now: what is relevant now, what works now and what doesn’t. However, there is another tradition in impact evaluation which is often overlooked or ignored, which is the historical approach. Every once in a while a historical evaluation is done (Jerve et al. 1999), and every once in a while somebody asks attention for this approach (van den Berg 2005), but it cannot be said to have been a strong tradition, nor a tradition that made a big impression. Complaints have been that historical evaluation studies are very expensive, are perhaps more research than evaluation, take a lot of time and are not impressive as regards learning, because lessons from years ago may not be relevant to the present circumstances, let alone the future (see for example the controversies surrounding the Dutch historical evaluations of long term relationships with several countries in van Beurden and Gewald 2004, pp. 63–67). So it is with some enthusiasm that the development community turned to experimental impact evaluation, preferably integrated into the design of projects and executed during their lifetime, and hoped that this would turn up relevant evidence of what works that would provide lessons for the immediate future. However, what if the evidence of what works and what doesn’t only reveals itself over time? What if the time horizon is in decades? What else are we to do but integrate historical approaches with other tools and methods?

Many problems in development are longer term in nature: to reduce absolute poverty, to reduce child-birth related death rates, to improve nutritional status, to integrate countries into the global economy, and so on – these are measured over decades and changes tend to happen relatively slowly. The Millennium Development Goals in general addressed global trends and impacts at higher scales. At these levels impact evidence can no longer be generated directly through experiments and other analytical tools such as meta-analysis, statistical analysis and modelling tend to take over. The Millennium Development Goals were monitored through statistical data. As 2014 report on the achievements of the Millennium Development Goals states: “reliable and robust data are critical for devising appropriate policies and interventions for the achievement of the MDGs and for holding governments and the international community accountable” (UN 2014, p. 6). However, especially when complex programmes and policies need to be improved, evaluations and research have to play their role, as they can provide answers to questions why a certain trend is occurring. For this reason the 2030 agenda for sustainable

---

<sup>1</sup>A movement that has its roots in evidence based medicine (see [https://en.wikipedia.org/wiki/Evidence-based\\_medicine](https://en.wikipedia.org/wiki/Evidence-based_medicine)) and has spread to education, international development and other areas, where its characteristics may differ in some aspects.

development includes evaluation as a follow-up and review principle for the agenda (UN 2015, p. 37).

Evaluations at higher scales and at the global level are not often done and are difficult to design, implement and report on. Many problems could be mentioned, such as reliability and comparability of data, external validity of evidence of causality, but a particular problem that raises its head in relation to impact evidence, is the problem whether evidence at local levels and lower scales translates into evidence at global levels, at higher scales and over longer time periods. The first chapter of this book has dealt with this issue in detail. In 2013 I argued that a “micro-macro paradox”, which points to successes at the micro level that seem not to be reflected in trends at the macro level, is particularly relevant to the linkage between environment and development and thus to sustainable development which aims to achieve a balance between society, the economy and the environment (van den Berg 2013, pp 41–43). Climate Change provides good examples for this. Many climate change related interventions are successful and achieve what they set out to do. However, the success of individual activities has not affected global climate change substantially. As the Intergovernmental Panel on Climate Change concluded in its 2014 report: “without additional mitigation efforts beyond those in place today, and even with adaptation, warming by the end of the twenty-first century will lead to high to very high risk of severe, widespread and irreversible impacts globally (*high confidence*)” (IPCC 2014, p. 77). Notice the use of the term impact for global phenomena.

### 3.2 Demand for Impact Evidence

Although the evidence movement has aimed to narrow down and reduce the meaning of the term “impact” as referring to what can be found through counterfactual testing, the term impact is an ordinary word in the English language, the meaning of which varies according to context. While science and in this case evaluation may prefer a precise definition and a narrow meaning of terminology, in general this will not change how terminology is used in conversation and debates. When the public demands to see proof of impact, they will use the term impact in an undefined way. To correct the public tends to be rather difficult if not impossible. The question thus emerges whether narrowing the definition of impact is helpful and whether another approach would not be more appropriate, which is to identify how the term is used, what kind of evidence would be required to meet the demand and to identify clearly what the advantages and disadvantages are of the tools and thus of the reliability, validity and credibility of the evidence.

A good example of the discrepancy between what works and does not work at the local level and whether “impact” is achieved according to the way the public thinks about it, is climate change. At the level of individual activities good, solid evidence is found on what works, especially on mitigation of climate change. Mitigation activities aim to reduce the level of greenhouse gas emissions and thus

aim to reduce the inflow of carbon dioxide and other greenhouse gases into the atmosphere. If a sufficient number of these activities take place, it should be possible to stabilize or even reduce the concentration of greenhouse gas molecules in the air, which is currently about 400 particles per million. While individual activities may be quite successful in reducing emissions, the overall concentration of greenhouse gases in the atmosphere continues to increase. There are thus two kinds of impacts that the public is concerned about: do individual interventions work and lower the emissions, and is climate change stopped? The first question may be answered through counterfactual experimentation, modelling or through before/after measurements of greenhouse gas emissions. Nothing about this is as simple as it sounds. The calculation and measurement of greenhouse gas emissions is not yet based on full understanding, agreement on principles and validation through international norms and standards. For an overview of the issues and what the current state of the art is, see STAP (2013).

All the successes of achieving impact at project level have so far not been able to change the overall trend in climate change, which is that the global mean temperature continues to rise. When asking for evidence of impact, donors and the public want to know whether projects have an impact, whether the project delivers and the causal mechanism that it embodies works. But donors and the public also want to know whether this leads to changes at higher levels, beyond the direct influence of the project, and ultimately they would like to see climate change stopped or even reversed. The demand for impact evidence is legitimate at all levels and cannot be met by referring to impact evidence only at project level or in the context of one intervention or one causal mechanism. Understanding the range of questions on impact evidence will enable evaluators to focus on the key questions that need to be asked in evaluations and will enable them to identify the tools and methods that need to be used.

### 3.3 Theories of Change for Climate Change Mitigation

The standard approach to identify key questions in an evaluation is to look for the “theory of change” that identifies how the intervention is expected to achieve impact. In traditional impact evaluations this leads to an identification of the causal mechanism that is supposed to “work”. In climate change, this is usually a combination of a technical mechanism and a behaviour mechanism: “if this new technology is adopted by people/institutions/countries it will lead to reduced greenhouse gas emissions and thus to a lower rate of global warming”. Traditional impact evaluations tend to focus on what works to effectuate behaviour change. If the behaviour change occurs, the intervention “works” and should be promoted. If it does not work, it should be stopped.

Organisations like 3ie, devoted to promoting traditional impact evaluations, are very much aware that this simple version may lead to all kinds of perverse effects that need to be taken into account or looked at, and for this reason they advocate that

impact evaluations should be “based on a thorough analysis of an intervention’s theory of change”<sup>2</sup> as there may be other links in the causal chains that should be tested or taken into account. Adopting the new technology is a change of behaviour, but it could potentially lead to unintended consequences which may lead to an overall increase of greenhouse gas emissions, if energy use increases overall. Other changes in the context may make a specific behaviour change redundant, as for example where new markets emerge and take over functions that are done more efficiently through new technology. However, the focus remains on checking for evidence of the behaviour change, as this is the causal mechanism that can be checked in a traditional impact evaluation. Let us explore whether a deeper understanding of the theory of change would lead to different and new questions.

Let us take a typical mitigation intervention as an example: the introduction of a new technology that would lower greenhouse gas emissions. The Hilly Hydel project in India was a typical project funded by the Global Environment Facility and the Government of India, supported through UNDP, which took place from 1995 to 2003. This has been a particularly well evaluated project (see Ratna Reddy et al. 2006). It was the object of a case study for a major GEF study on local benefits generated through support for global benefits (GEFEO 2006), has an end-of-project evaluation including a counterfactual impact assessment (Ittyerah et al. 2005) and was further studied for the GEF impact evaluation of mitigation projects in emerging economies (GEFIEO 2013). For a total amount of \$ 14.6 million this project led to the introduction of small hydroelectrical power plants in hilly regions in India, mostly in remote villages without access to the main grid. The reduction of greenhouse gas emissions was supposed to be achieved through using a renewable source of energy (hydro power) and reducing the need for wood as a source for fuel, thus leading to a secondary but important benefit: reduced deforestation. The outputs of the project were a national strategy and master plan for hydro electrical power generation, 20 stand-alone small hydel power generating water mills, upgrading of 100 existing water mills to incorporate power generation and institutional and human capacities to ensure sustainability. In general these outputs were achieved or surpassed – upgrading of no less than 143 water mills took place. All in all this led to direct greenhouse gas emission reductions of 1900 tons CO<sub>2</sub> equivalent per year. If the potential for installation of these small-scale hydroelectric water mills would be fulfilled throughout India, the total amount of reductions per year would calculate as 4 million tons CO<sub>2</sub> per year (GEFIEO 2013, table 24 p. 70).

The theory of change of the project focused on introducing a technology that was new for the villages in the hilly areas, that would lead to a source of energy that would be more reliable and would lead to a halt to deforestation because of energy needs, reduced greenhouse gas emissions as a result and given its benefits, would convince villages to invest in this kind of technology. This would lead to a change in the market for rural energy in hilly areas, where hydroelectricity would take the

---

<sup>2</sup>See From Influence to Impact. 3ie strategy 2014–2016, p. 2, found at [http://www.3ieimpact.org/media/filer\\_public/2014/09/07/3ie\\_strategy\\_summary\\_final\\_rgb.pdf](http://www.3ieimpact.org/media/filer_public/2014/09/07/3ie_strategy_summary_final_rgb.pdf), on September 4, 2015.

place of wood burning and fossil fuel generators, also resulting in less pollution in these villages. The behavioural assumption was that villagers would be willing to spend more money on energy given the benefits in reliability of supply, reducing the need for wood and thus reducing deforestation, reducing pollution and saving time in searching for wood. The hydroelectric power plants would be made available through public-private arrangements, supported by the States and by the Federal Government, and legitimized and promoted through a national strategy. The theory of change provided a series of causal linkages that together would change the market for hydro-electric power in remote hilly areas and would lead to considerable reductions of greenhouse gas emissions. More challenging was the perspective that this would also lead to reduced deforestation and to biodiversity benefits (Ratna Reddy et al. 2006, 4071).

The demand for evidence of impact can be placed at various levels in this theory of change. First of all, the hydroelectric power plants are supposed to produce energy with greater efficiency in greenhouse gas emissions than other local energy sources: these emissions should be lower than the same levels of energy produced through burning wood and through fossil fuel generators. Technological expectations in this regard need to be met and one could argue that the first impact question would be whether the hydroelectric plants deliver what they promise. The second question is whether the village manages to integrate the hydroelectric mills into their society: will they maintain the mills, pay their energy bills and use this source of energy instead of reverting to wood and fossil fuels? This is the kind of behaviour question that is beloved in traditional impact evaluations. A third question concerns whether the shift towards hydro-electric power is leading to a change on the energy market in remote hilly areas. Have demonstration and the first verifiable outputs of the project led to an increased supply on this market; i.e. is the private sector offering hydroelectric technologies to villages? And if so, is there a demand for this? Are villages actively taking this up for consideration when looking at their energy options? And is the financial sector willing to provide loans for investments to the communities or villages? A fourth impact question is then whether the market has changed – if it has changed – locally, regionally or nationally. These questions need to be looked at from three different perspectives: time, space and scale.

### **3.4 Key Questions Related to Time, Space and Scale**

Especially with a global issue like climate change the demand for impact evidence ranges from “what works here and now” to “has it contributed, or will it contribute, to stop climate change”. The first is very local, time and scale bound, just looking at whether a specific mechanism works as it is supposed to. The second looks at the planet, at scenarios that go into the future and that are at the highest (global) scale. Both are relevant questions and need to be answered.

This translates into issues of time, space and scale. It is quite clear that a project of \$ 14.6 million cannot change the national energy market for remote hilly areas

overnight. This takes time; in fact the impact assessment done at the end of the project asked for “adequate time” to pass and for a stable situation to be achieved before impact is assessed (Ittyerah et al. 2005, p. xv). And if individual projects need adequate time to have an impact, it follows that market change can only be observed and measured over even longer stretches of time. Longer time lapses are well known in environmental circles and on environmental impact, as Hildén (2009) and Rowe (2014, 54–55) have pointed out, but they tend to be less associated with market change. The slow pace of market change is more often observed with impatience, raising the question why no change is happening, which led Wörlén (2014) in her study of climate mitigation evaluations to reformulate the “theory of change” approach to a “theory of no change approach” that focuses on a better understanding of market barriers and how they can be overcome.

In general environmental boundaries do not follow jurisdictional boundaries. One ecosystem may spread over several countries, and one country may have several ecosystems. Rowe (2012) asked attention for the fact that location may differ conceptually and practically between a social and economic system that is targeted for change and an ecosystem that is influenced through the same intervention or action. But this is not only an issue of different locations of systems, but also of scope of an intervention: it may be focused on a direct impact in the villages in which it is implemented, while other areas are still outside the scope of the project or have not yet been approached by suppliers, or invited to participate by State or Federal government.

It is an issue of scale when impact needs to be observed at several levels: that of energy supply and demand, of greenhouse gas emissions related to energy, of greenhouse gas emissions including deforestation and alternative sources of energy, of livelihood and financial resources issues in the villages, of hilly rural areas in general, and perhaps somewhat more removed, whether greenhouse gas emissions in India are positively influenced by what happens in remote hilly areas. The last does not seem likely, and it may lead to a feeling of disenchantment – if it does not help India, it does not help the world, and it does not stop climate change.<sup>3</sup> But that was the reason the project was co-funded by the Global Environment Facility in the first place!

Scale is not easily defined. It seems clear that while interventions or actions move from one actor to multiple, from one location to many, from a “local” to a “national” or even “global” level that moving up scales is involved, but scales can also be understood in terms of different dimensions or sectors. Kennedy et al. (2009) recognises jurisdictional and management dimensions as different scales, and Bruyninckx (2009) asks attention for overlap and discrepancies between social, economic, environmental and spatial scales. Yet even though there is no universal agreement on how scales should be defined or what their boundaries are,

---

<sup>3</sup>And a good overall conclusion on the project was formulated by Ratna Reddy et al. (2006, 4078): the overall impact of the project appears to be slightly positive or neutral in a majority of key indicators. Certainly not a major contribution to reduced greenhouse gas emissions as hoped for.

there is widespread agreement that to mainstream, replicate, reproduce, upgrade or upscale interventions to higher levels is an essential perspective in understanding causal pathways from the micro-level to higher level goals.

Garcia and Zazueta (2015) argue that at higher scales interventions should be interpreted and looked at from a systems perspective. Individual components and elements do not a system make, but when they start interacting, they tend to take on characteristics of a system, which can have its own dynamics and shifts and changes. Arguably markets operate as systems and market change is systemic change: subtle changes in supply, demand and enabling environment can lead to “tipping points”, after which slow, reversible change becomes irreversible, or the point in time at which a new technology (such as hydel power) becomes mainstream.

In conclusion key questions related to **time** lead to the realisation that impact can be measured at each moment in time – ex ante as impact assessment, through modelling and calculations, real time through monitoring, experimental design, trend analysis etc. and ex post through various evaluations and studies. Key questions related to **space** make us realise that impact differs per area and that areas have different impacts. Key questions related to **scale** point to the need to mainstream, replicate, upscale and broaden the scope of interventions before impact can be achieved at higher levels.

### 3.5 Using Time and Space to Identify Approaches

In principle the three dimensions of time, space and scale can be used to build a three dimensional matrix in which the theory of change of an intervention, programme or policy can be represented. This will enable the evaluator to identify where a particular demand for impact evidence needs to be placed, and what would be appropriate analytical tools to evaluate impact. Figure 3.1 presents a matrix of time and space aspects. The time dimension goes from ex ante (designing and formulating a new intervention) to important moments in real time (from inception to mid-term to end-of-project) to ex post and identifies ex post evaluation approaches. Red “balloons” signify evaluation approaches; blue ones monitoring and data analysis, whereas a green balloon identifies a research approach. Of course evaluations use and analyse monitoring data, and often use research tools and methods. Figure 3.1 just presents a possible configuration of what is dominant in the matrix from an evaluation perspective. The space dimension goes from local through national and regional to global, but has an extra row for ecosystems, which overlap with other rows.

The ex ante column is occupied by ex ante evaluation and impact assessment, which is a lively community of practice that uses various methods and tools to come to conclusions on the potential impact that different scenarios may have throughout time. These impact assessments tend to use modelling as their preferred tool and may present several scenarios that would lead to different impacts. The ex post



	Ex ante	Inception	Mid-term	End of project	Ex post < 2 years	Ex post 5-8 years			
Local	Experimentation								
National	Ex ante evaluation and impact assessment	Monitoring and data analysis (including "big data")				Terminal evaluation	Mixed methods ex post impact evaluation		
Regional									
Global									
Ecosystem (overlap with other rows)					Natural experimentation				

**Fig. 3.1** The time and space dimensions of demand for impact evidence (Source: Author)

evaluation community tends to keep its distance from the ex ante evaluators, as there is widespread concern that any involvement of ex post evaluators in ex ante evaluation will lead to a conflict of interest when the activity needs to be evaluated later on. If design and implementation characteristics were decided upon because of an ex ante evaluation’s outcomes, an ex post evaluator would in fact be required to evaluate his or her own judgments in the ex ante evaluation. In actual practice the two communities of practice hardly mingle. Ex ante evaluators have their own conferences and their own literature and good practice standards. What Fig. 3.1 shows is that they are the first to delve into the question of impact and aim to provide evidence, even if hypothetical at that stage, for what an intervention would set out to do.

During implementation monitoring and evaluation often become management tools. If the project needs to be steered through difficult circumstances and react adequately to changes, it needs to set up an adequate monitoring system, either collecting its own data or using data from available statistical services. Relatively new is the inclusion of real time evaluation, which on impact tends to take the form of randomized controlled trials that need to be included in the design of the project and need to be adhered to during implementation, in order to come to valid conclusions about the causal mechanism tested out. Other evaluations during implementation (such as mid-term evaluations) tend to look at processes and efficiency and are not represented in this matrix. Randomized controlled trials tend to be “local” in nature; rarely will we see RCTs at the national level and even more rarely at the regional level, as they would become very costly to reach a sufficient level of data (large “n”) to allow for conclusions at that level.

In the ex post columns we tend to see two varieties of evaluation that provide impact evidence. First of all, end-of-project evaluations may present results of experiments or provide data on impact; usually these evaluations also contain

important information on expected “progress toward impact” (van den Berg 2005) and whether the conditions have been set to enable longer term impact. The last column presents ex post evaluations 5–8 years after the project has ended. These are almost invariable historical evaluations, using a historical approach to trace whether the results of the project have contributed to observed changes in trends, markets, societies, economies and the environment. These evaluations tend to advertise themselves as “theory of change” oriented and using mixed methods and triangulation of evidence to come to conclusions on impact. They have less of a problem to move beyond countries to regions and the global level, not because the evidence is stronger after 5–8 years, but because they are more flexible in approach and are more pragmatic and adaptable in using data sources and linking data where possible. This sounds opportunistic, but there are many scientifically sound methods and tools that can be combined and triangulated, as amply demonstrated by Stern et al. (2012) and Garcia and Zazueta (2015).

### 3.6 Using Time and Scale to Identify Approaches

Another cut-through of the three-dimensional matrix of time, space and scale would be to combine time and scale. Figure 3.2 presents this matrix. The time dimension is of course the same as in the time-space matrix, but has been simplified a bit, for example presenting one row for ex-post rather than two. The scale dimension provides various perspectives of scale. From interventions focused on one causal mechanism, such as a project focusing on changing customer behaviour on the energy market through price setting, to multiple interventions within one project, of for example public-private partnerships, social change movements, capacity development efforts, to a perspective on an enabling environment that through rules and regulations, taxation, knowledge dissemination and other incentives tries to redirect a market or change behaviour, to market change and transformation, the interventions become more complex and challenging to evaluate. At the far right I have included climate change, and again this environmental scale overlaps with others, posing a special problem that two evaluands need to be recognized in an evaluation that includes environmental objectives (see Rowe 2012).

Again we see randomized controlled trials and quasi-experimental approaches focusing mostly on one intervention, as to control for combinations of interventions will become very costly. Ex ante research will deliver counterfactual assessments of how different scenarios will perform at all scales. A relatively new method such as Qualitative Comparative Analysis (QCA) is currently often used for case studies of more complex interventions and the enabling environment. Markets are of course the subject of economic research and for evaluations especially market research to assess whether a new product or approach has a chance on the market dominates in the market columns and ideally before the new intervention starts. At the programme and policy levels, ex post impact evaluations may look at triangulation

	One intervention	Multiple interventions	Enabling environment	Market change	Market transformation	Climate change (overlap with other rows)
Ex ante	Counterfactual assessments			Market research		
Inception	Before/after,					GHG data
Mid-term	RCTs and quasi-exp.					
End of Project		QCA		Economic modelling	Triangulation, mixed methods	
Ex post						

Fig. 3.2 The time and scale dimensions of demand for impact evidence (Source: Author)

of different sources of evidence (including monitoring data on Greenhouse Gas emissions) and use a mixed methods approach (as advocated by Bamberger 2012).

### 3.7 Using Space and Scale to Identify Approaches

When the two dimensional cut of space and scale are taken out the three dimensional time, space and scale matrix, at first sight a less well covered picture emerges, with some clear gaps where currently no favourite tool or method for evaluation seems to be in use. Figure 3.3 presents the rows used in Fig. 3.1 with the scales used in Fig. 3.2. I have focused methods and tools on what they are mostly used for and where their recognized strength is. Randomized controlled trials dominate in providing impact evidence on local interventions that focus on one causal mechanism. When we are looking at multiple causal mechanisms and interventions moving beyond national boundaries to regional collaborations, quasi-experimental methods and QCA become more or less dominant. Social network analysis is a particularly powerful analytical tool that could help in complex interventions with many partners, including the enabling environment that supports actors in participating in societal or economic action. The Delphi methodology has been used to evaluate market change, as market experts may be able to identify why changes have occurred and what would have happened without changes in the enabling environment or if certain technologies would not have become available. Research methods such as modelling take over on the right side of the matrix. The gap in the lower left hand corner of the matrix could potentially be an expression of costs: it would be prohibitively expensive to do global or ecosystem wide randomized controlled trials, while theory of change oriented

	One inter- vention	Multiple inter- ventions	Enabling environ- ment	Market change	Market transform- ation	Climate change
Local	RCTs	Quasi- experimental and QCA	Social Network analysis	Delphi	Modelling of data and experimentation (quasi- and natural)	
National						
Regional						
Global						
Ecosystem (overlap with other rows)						

**Fig. 3.3** The space and scale dimensions of demand for impact evidence (Source: Author)

mixed methods evaluations may see it as a waste of money to focus on one causal mechanism only.

Potentially meta-evaluations and meta-analysis could go a long way towards covering some of the gaps, as has been advocated by the evidence movement through so-called systematic reviews. However, there are methodological problems with these reviews. They tend to focus on a specific question and go through a huge number of studies and evaluations to see whether they provide evidence on that specific question. Many studies turn out not to have evidence for that question and thus are not used. Another issue is that these systematic reviews tend to not accept evidence that is gathered outside the narrow range of methods that are considered by the evidence movement to be sufficiently rigorous.<sup>4</sup> More recently realist perspectives have started to become more fashionable in meta-evaluations, which broadens the range of evidence that is accepted. An example can be found in Chap. 13, ‘What do evaluations tell us about climate change adaptation’ of this book.

### 3.8 Conclusions

There is a famous scene in the British comedy Fawlty Towers which provides a good metaphor of how impact evidence may be treated by a narrow interpretation of evidence based politics. In this particular episode of Fawlty Towers the hotel manager Basil Fawlty puts some money on a horse in the hope of substantial

<sup>4</sup>See for example <http://www.whatworksgrowth.org/resources/the-scientific-maryland-scale/>. See also the discussion in [https://en.wikipedia.org/wiki/Hierarchy\\_of\\_evidence](https://en.wikipedia.org/wiki/Hierarchy_of_evidence).

earnings, and desperately wants to keep this secret from his wife Sybil. But the Spanish waiter in the hotel, Manuel, discovers what goes on. Basil asks Manuel to deny, if Sybil would question him, that he has any knowledge of this. When Basil is discovered by Sybil in suspicious circumstances with a lot of money, he needs to prove that he came by this money through legal means, and he asks Manuel to vouch for him. Manuel looks at Basil, grins, and in a proud performance exclaims: “I know nothing”. After a few seconds he repeats, with added emphasis: “I know nothing”, thus sealing Basil’s fate. The evidence based movement came to the foreground and argued for randomized controlled trials and counterfactual impact evaluations by claiming that old fashioned evaluations could be thrown in the wastepaper basket, and that there was a serious gap in evidence that needed to be filled. On international cooperation the evidence on what works and what doesn’t was, to adapt Manuel’s phrase: “we know nothing”. However, an analysis of the dimensions of time, space and scale demonstrate that randomized controlled trials are particularly good at covering a few of them, and that in many cases evaluators will need to explore other methods and tools to provide evidence on impact. As a result of the narrow scope of evidence that is accepted by the evidence movement, they will have difficulty in explaining to policymakers, boards and parliaments that what they want to see evidence on cannot be provided through randomized controlled trials.

The three dimensional matrix of time, space and scale provides a systemic ordering of demand for impact evidence, and inspiration for how this can be uncovered through various evaluation techniques. It underscores the wide range of scientific tools and approaches as discussed in the Stern report (2012). Further analysis is needed. No doubt more scientific tools exist and can be placed in the matrix. It could be developed as a heuristic tool to identify key evaluation questions and approaches. It also demonstrates that impact evidence is available throughout the cycle of projects, programmes and policies and that demand for impact evidence can be throughout the lifetime of a project and will get to higher levels and scales after the project has ended.

In the case of climate change mitigation, the matrix provides a better understanding why impact is visible at project level and in markets directly influenced (and hopefully changed) by the project, but that impact at the global level is illusive, not visible, and has not led to the desired change in trends. Especially where goals are formulated at the highest level the matrix may be useful in providing a systematic understanding why impact cannot (yet) be demonstrated at that level.

My suggestion is to further develop the matrix as an analytical tool to:

1. Better identify the demand for impact evidence: is it on whether a specific causal mechanism works, or is it whether the problem that needs to be addressed is becoming solved, or whether global, regional or national trends are moving in the right direction, and if so, how that is linked to the intervention.
2. When the demand is identified, how would this translate to key evaluation questions that focus on the right moment in time, at the right location and at the appropriate scale?

3. Given these questions, the appropriate evaluation approaches and tools and methods can be found to address them.
4. Lastly, by framing the evidence in time, space and scale the evaluation can better explain why evidence is generated in the way that is chosen, and why other methods (such as randomized controlled trials in the case of complex interventions, or mixed methods case studies in the case of a straight-forward intervention that is localized and focuses on testing one causal mechanism).

The Centre for Development Impact in Brighton will continue to work on this tool and aims to further develop it along these lines.

## References

- Bamberger, M. (2012). *Introduction to mixed methods in impact evaluation*. InterAction [etc.] [Impact Evaluation Notes: No.3 August 2012].
- Bruyninckx, H. (2009). Environmental evaluation practices and the issue of scale. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental program and policy evaluation. New Directions for Evaluation*, 122, 31–39.
- Garcia, J. R., & Zazueta, A. (2015). Going beyond mixed methods to mixed approaches: A systems perspective for asking the right questions. *IDS Bulletin*, 46(1), 30–43.
- GEFEO (2006). *The role of local benefits in global environmental programs*. Washington, DC: Evaluation Office, Global Environment Facility.
- GEFIO (2013). *Climate change mitigation impact evaluation. GEF support to market change in China, India, Mexico and Russia*. Washington, DC: Independent Evaluation Office, Global Environment Facility [Unedited version, downloaded from [gefio.org](http://gefio.org) in August 2015].
- Hildén, M. (2009). Time horizons in evaluating environmental policies. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental program and policy evaluation: Addressing methodological challenges. New Directions for Evaluation*, 122, 9–18.
- IPCC. (2014). Climate change 2014: Synthesis report. Contributions of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. Geneva: IPCC.
- Ittyerah, A. C., Choudhary, R., Narang, S., Choudhary, S. D. (2005). *Terminal evaluation and impact assessment of the UNDP/GEF project – IND/91/G-31 – optimizing development of small Hydel Resources in the Hilly Regions of India*. S.I., s.n.
- Jerve, A. M., et al. (1999). *A leap of faith: A story of Swedish aid and paper production in Vietnam – the Bai Bang project, 1969–1996*. Stockholm: SIDA.
- Kennedy, E. T., Balasubramanian, H., & Crosse, W. E. M. (2009). Issues of scale and monitoring status and trends in biodiversity. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental program and policy evaluation: Addressing methodological challenges. New Directions for Evaluation*, 122, 41–51.
- Ratna Reddy, V., Uitto, J. I., Frans, D. R., & Matin, N. (2006). Achieving global environmental benefits through local development of clean energy? The case of small hilly hidel in India. *Energy Policy*, 34, 4069–4080.
- Rowe, A. (2012). Evaluation of natural resource interventions. *American Journal of Evaluation*, 33, 384.
- Rowe, A. (2014). Evaluation at the nexus: Principles for evaluating sustainable development interventions. In J. I. Uitto (Ed.), *Evaluating environment in international development*. London: Routledge.

- STAP. (2013). *Calculating greenhouse gas benefits of the global environment facility energy efficiency projects, version 1.0*. s.l.: Science and Technical Advisory Panel (STAP), March 2013.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations* (Working paper 38). London: DFID.
- United Nations. (2014). *The millennium development goals report 2014*. New York: United Nations.
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. New York: United Nations [A/Res/70/1], downloaded from <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication> on 3 Sept 2015.
- van den Berg, R. D. (2005). Results evaluation and impact assessment in development co-operation. *Evaluation*, 11(1), 27–36.
- van den Berg, R. D. (2013). Evaluation in the context of global public goods. In R. C. Rist, M.-H. Boily, & F. Martin (Eds.), *Development evaluation in turbulent times: Dealing with crises that endanger our future*. Washington, DC: The World Bank.
- van Beurden, J., & Gewald, J.-B. (2004). *From output to outcome? 25 years of IOB evaluations*. Amsterdam: Aksant.
- Wörten, C. (2014). Meta-evaluation of climate mitigation evaluations. In J. I. Uitto (Ed.), *Evaluating environment in international development* (pp. 87–104). London: Routledge.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

