

12 Theories of change and the evaluation of sustainable impact

Moving beyond simplicity in development cooperation

Peter van der Knaap

Introduction

Theory-based evaluation offers an appealing approach for both evaluators and policymakers. This seems to hold true, especially in the field of development cooperation, where theories of change have long been an important part of the modus operandi. It is no surprise: policy theories provide focus for policymakers and allow for a logical starting point for important evaluation questions, such as: Did the intervention work? and Which improvements are possible?

Theory-based evaluation has also been the subject of fierce criticism, mainly because of its tendency to simplify societal challenges, public policy programmes, and the purpose of evaluation. With the adoption of the UN Sustainable Development Goals, this criticism is more relevant than ever: like other policy areas, development cooperation is shifting from individual projects to complex programmes and change on a national system level.

So, what does this rise of sustainability mean for the lure and relevance of theories of change and theory-based evaluation? This contribution starts from the still-strong promise of theory-based evaluation by revisiting its strengths and limitations (section 1). In section 2, the new millennium imperative that development policy must be sustainable is discussed. As policymakers are moving from simple projects to complex development interventions that aim to change entire social systems, evaluators are faced with growing complexity. Section 3 discusses alternative evaluation methods to evaluate effectiveness and makes the case for a new system-based theory of change approach. Herein, a theory of change not only serves to express the rather straightforward relationship between ‘measure’ and ‘effect’, but it also specifies the necessary system requirements for sustainable impact.

Section 4 concludes with a discussion of future directions of the evaluation of sustainable impact. Four avenues are explored: (1) the quest for sustainable development may be a new lifeline for the useful employment of theories-based evaluation by including system-level ‘theories of sustainable change’; (2) there are useful, context-driven ‘back tracing’ and responsive methods available for system change or ‘transition’ evaluations; (3) the ‘traditional’ evaluation of

individual policy measures still has great value; and (4) when seeking sustainable development, policymakers should be beware of the risk of ‘overreaching’.

1. Theory-based evaluation revisited

Decades after its inception, theory-based evaluation remains promising to both scholars and practitioners. In essence, it can be defined as the analysis and valuation of the contribution of a theory of change, ‘intervention logic’, or ‘policy theory’ to resolving or controlling society’s problems. Herein, an intervention strategy is placed in rank and order to achieve goals—perfect for simple policy interventions.

Despite three decades of postmodernism, increasing policy complexity, and the rise of populism in Western democracies, policy and evaluation are still regarded as strong rational-analytic concepts. The rationale for public policy measures and, hence, public expenditure is to work systematically and even scientifically to achieve democratically legitimized goals. The usefulness of theories of change in evaluation and monitoring is present in all four phases of the policy cycle: (1) policy development, (2) debate and decision making, (3) implementation, and (4) rendering account. In all phases, policy-oriented learning and *focus* are important—and using theories of change provides just that (Van der Knaap, 2003).

To give an example, the theory of a simple programme may be: ‘If we vaccinate children and provide clean drinking water and sanitation, then less children will die of disease before the age of five’ (compare with USAid, 2022). Evaluation *ex ante* focuses on the expected causality and cost effectiveness of measures. In this way, using theories of change may enable better-informed, evidence-based decisions on policy proposals. Furthermore, a theory of change can be translated into an intervention-logic or implementation scheme, in which resources, actions, outputs, and results are combined. This allows for planning, monitoring, and management of implementation processes and progress but also for learning, including fundamental lessons about the validity of the theory itself.

Theory-based evaluation *ex post* revolves firstly around the assessment of the extent to which policy has resulted in success: Did the intervention work? Did it help to achieve the intended effect? Taking theories of change as a starting focal point for (summative) evaluation *ex post* creates a strong framework of reference and can boost the setup, implementation, delivery, and utilization of evaluation research. The second goal of theory-based evaluation is a better understanding of the underlying causal mechanisms. In such a case, the central question is: Were the assumptions on which the policy programme was based ‘right’ or not, and why? In this way, theories of change will serve both the notion of accountability and learning.

But like all theories, any theory of change is an abstraction of reality and will inevitably exclude a great wealth of facts and ideas—and deliberately so. In this

way, people get focus. The result, however, may very well be a simplified version of reality. Moreover, every such theory—and especially those that are well articulated—will direct perception, interpretation, and (thus) assessment.

It is difficult to underestimate the potential simplifying effect of a theory of change and the effects thereof on decision making, monitoring, and evaluation. Where too much emphasis is placed on anticipated causality and measurability, the result may be tunnel vision, rigidity, and even fear of and resistance to innovation (Van der Knaap, 2003). Everything that cannot be expressed in theories, performance data, and objectives may escape the attention of both policymakers and evaluators.

2. Sustainable development: From simple programmes to system-level complexity

In my 1988 student edition of *The Concise Oxford Dictionary* the word *sustainable* is not listed. *Sustain* is defined as ‘hold up’ and ‘keep from failing or sinking, especially for a prolonged period’, but the adjective itself is, rather surprisingly, lacking. Since the 1990s, however, *sustainable* has taken off in a spectacular way. In 2022, efforts to do good must lead to lasting results: road safety professionals to realize ‘sustainable road safety’ and innovation policy programmes to shift entire sectors into new products or services. This fundamentally changes the way policymakers and evaluators ought to consider ‘effect’, ‘effectiveness’, and ‘impact’.

Let’s take development cooperation as an example. Herein, the UN Sustainable Development Goals constitute ‘bold and transformative steps... to shift the world onto a sustainable and resilient path’. It is no surprise that, in its new 2021 evaluation criteria, the Organization for Economic Cooperation and Development (OECD) included sustainability as a new evaluation criterion for development efforts. The main question perfectly summarizes the key point: Will the benefits of a policy intervention last? What is the extent to which the net benefits of the intervention continue, or are likely to continue? (OECD, 2021). According to the OECD, to assess sustainability, the evaluator must examine ‘the financial, economic, social, environmental, and institutional capacities of *the systems needed to sustain net benefits over time*’ (OECD, 2021, emphasis added).

In short, the rise of sustainability means the evaluation must consider resilient impact. This leads to ‘tougher’ questions, such as: Did we succeed in establishing a better ‘enabling ecosystem’ for the benefit we seek? or Will the benefit last when the policy programme is over?

The difference between implementing and evaluating a simple project (like a vaccination project) and a complex development intervention (like improving an entire country’s health system for the future) is enormous. Simple projects work according to ‘blueprint’ approaches, producing standardized outputs with

relatively linear causal relationships between output and defined objectives (outcomes), within a specific timeframe (see Figure 12.1). In addition, agencies usually deliver a limited number of services to a defined and often small target population.

The Sustainable Development Goals' imperative that development and development measures must be sustainable—and therefore aim to change the underlying systems needed to eradicate poverty, secure human rights, and achieve climate objectives—means that policymakers and evaluators can no longer limit themselves to simple programmes but must aspire to (also) undertake complex development interventions.

As the OECD puts it: When the aim is sustainable change, the evaluation must take a broader perspective than individual measures and consider more fundamental, enabling changes in political, economic, financial, social, and environmental systems (OECD, 2021). In other words: evaluation must allow for complexity.

3. Alternative evaluation methods: Allowing for complexity

There is growing consensus that policymakers and evaluators alike must recognize the complexity of human behavior and the government and societal systems we create to live our lives (see Essay 11 by Houlberg and Rieper). There are vast numbers of individual elements at work that interact in complicated ways, whether it is ecosystems, energy networks, financial markets, or phenomena such as urbanization and migration (OECD, 2017).

The challenge of designing and applying the theory of change in such complex contexts is pronounced in many essays in this book including those by Morkel (Essay 14) and Lima & Lafer (Essay 13). Assessing outcomes of complex interventions poses special challenges for the evaluator (compare with USAid, 2022):

- Complex policy programmes may lack clearly defined activities and/or objectives and/or timelines.
- There are often multiple actors responsible for delivering activities and outputs.
- As interventions have a long-term horizon and a regional or countrywide reach, baselines and credible counterfactuals are often lacking.
- Even when evaluations include a counterfactual, the causes of observed changes may not be fully explained.



Figure 12.1 Model of a simple project

In evaluation literature, this has led to the development of alternative, more explorative and responsive methods that do not explicitly start off with a theory of change (Van der Knaap, 2011; compare with Essay 3 by Palenberg). Rather than the rational analytic ‘golden standard’ of randomized controlled trials (RCTs) or statistical (regression) analyses, the appraisal of causality uses a more constructivist approach. In addition, the importance of context is stressed. Examples of evaluation models that connect policy programmes to outcomes—or often vice versa—are outcome harvesting, *modus operandi*, realist evaluation, and contribution analysis:

- *Outcome harvesting* has specifically been developed to evaluate policy interventions in complex and dynamic contexts in which no simple theories of change can be constructed (compare with Wilson-Grau & Britt, 2013). The essence is that effects are not evaluated against objectives in policy programmes, but rather that the evaluator ‘harvests’ any change that has occurred in the domain of the programme. Once changes have been identified, an attempt will be made to trace the effects back to the measures undertaken under the programme. Lessons can be drawn on combinations of measures and conditions that ‘work’.
- A similar approach can be found in the so-called *modus operandi* model (Scriven, 1974). Here, too, the objective is to—backwardly—infer causality between outcomes and programmes. Drawing from forensic science, the evaluator makes a list of possible causes for results and looks for evidence in a bottom-up fashion. Again, the aim is to draw conclusions on combinations or modes that produce results.
- The premise of *realist(ic) evaluation* is that evaluation is useful when it provides for answers to the question ‘What works in which circumstances and for whom?’, rather than merely ‘Does it work?’ (Pawson & Tilley, 1997). The evaluator seeks to identify the underlying mechanisms that explain ‘how’ outcomes were caused and the influence of context thereupon (compare with Douthwaite et al., 2017).
- Differently, *contribution analysis* starts with an articulation of the theory of change, but this may be complex rather than simple. Instead of RCTs, multiple lines of evidence are included to consider causal links in a concluding ‘performance story’ about what needs to be done in terms of conditions and interventions to achieve results.

The main advantage of these methods in ‘the new age of sustainability’ is that they make it possible to move beyond simplicity through the combination of methodological approaches that support the investigation and assessment of complex policy programmes in diverse circumstances. In all this, we must guard against pride. Policy, however well-intentioned, cannot do everything. In addition, it often remains useful to evaluate individual measures for

effects—as part of a comprehensive programme ‘at system level’ but also on its own.

4. Towards system-level ‘theories of sustainable change’

As policymakers are moving from simple programmes to complex, system-level development interventions, it is time to ensure that evaluation does justice to increased complexity and system-level dynamics. Moreover, we must move towards a ‘system theory-based evaluation’ that embeds system approaches.

Herein, system change theories or *theories of sustainable change* are needed. Ideally, in the field of development cooperation, these higher level, more encompassing theories will make it possible to design policy programmes that will be ‘self-supporting’. This means that all relevant beneficiaries buy into other coherent policy programmes and sustaining investments. The difference between simple intervention approaches and systems approaches is that this type of ‘maintenance’ is worked out and committed to in advance. In this way, results can be achieved for years to come.

This is important: as the OECD (2021) stresses, benefits often fade out after some time due to lack of maintenance and replenishment of new roads, bus fleets, teaching materials, hospital staff, and others. Systemic planning is necessary to have engineers trained, mechanics qualified, curriculum development experts and institutions in place, and so on. According to the OECD, ‘The role of evaluation here can be to scrutinize assumptions in the theory of change for how sustainability is achieved’ (2021, after Mansuri & Rao, 2012, p. 23; White et al., 2018, p. 24).

A theory of sustainable change not only serves to express the rather straightforward relationship between ‘measure’ and ‘effect’, but it also specifies the necessary underlying system requirements for sustainable impact. As in simple programmes, pathways can still be identified to connect these requirements to concrete conditions and actions. The difference is that there will be more than one relationship and that some will be conditional for others.

An example is the Expanded Program on Immunization, as described by Decouttere et al. (2021). Based upon a literature review, they constructed a ‘conceptual diagram’ in which the systemic objective to create ‘an environment for health system development’ is connected to the eventual impact of immunization through vaccination (see Figure 12.2).

A theory of change allows for focus but a system change theory or theory of sustainable change is undeniably more complex than a simple programme theory of change. Yet, without paying attention to underlying system requirements for sustainable impact, there can be no learning of lessons on the (likely) continuation of a policy intervention’s benefits in the medium to longer term. These lessons may also highlight the potential scalability of the sustainability measures of the intervention within the current context or the potential replicability in other

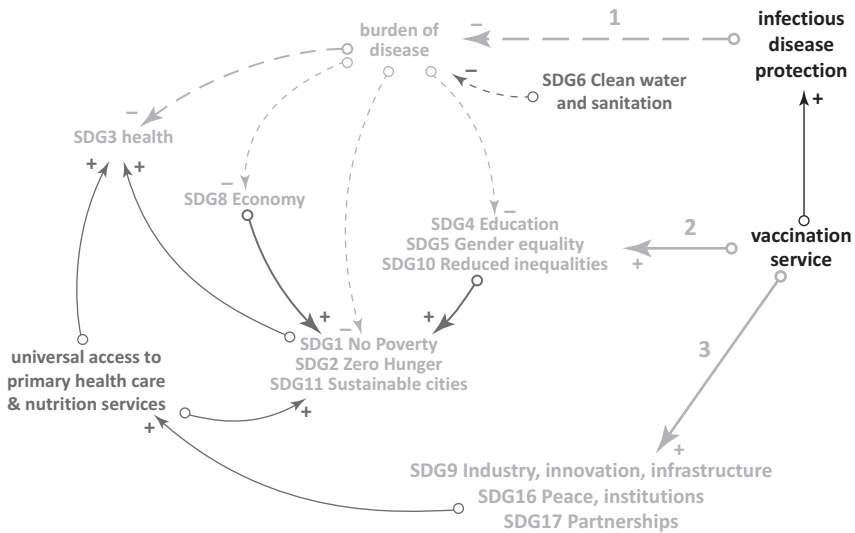


Figure 12.2 Example of a ‘conceptual diagram’ for the Expanded Program in Immunization

contexts (OECD, 2021). They should also be used in making decisions about discontinuation of programmes or aid: Is there an appropriate exit strategy that can lead to a continuation of positive effects, or not?

As I see it, policymakers and evaluators that seek to make a sustainable impact must adopt this broader perspective: efforts to improve health, safety, or our climate should no longer be considered ‘standalones’.

Both ex ante and ex post evaluation must examine the conditions for sustainable results. Ex ante, the question whether the necessary conditions for sustainable results can be met must be asked. Ex post, the first question is: Were those conditions considered at the start of a policy programme and incorporated in the design of the intervention and which lessons were learned along the way? The second question logically is: Are the necessary conditions for sustainable results in place and, hence, what can be said about the likelihood that investments in development will yield the desired results in the medium to longer term?

Four future directions

Putting the notion of sustainability central in policymaking and evaluation is crucial if we want to achieve and evaluate the effects, impact, and value-for-money of policy programmes. After all, what is money spent worth when the net result will be gone in just a few years after the intervention? For this reason, policymakers

and evaluators alike must take a system perspective, where a broader perspective is inherent to include the necessary conditions for sustainable results.

Despite the move from simple programmes to complexity, the promise of theory-based evaluation is still strong. True, a system perspective means more complexity, whereas one of the main advantages of theories of change and theory-based evaluation was to reduce complexity and provide focus. But even more sophisticated system change theories will enable policymakers to pay attention to the most important underlying causes and requirements for long-term successes. Even relatively complex theories of sustainable change can still provide a reference frame to stakeholders and other parties involved in the policy-oriented debate, allowing for well-considered argumentation.

These functions remain relevant to evaluators. Opting for the ‘official’ theory of sustainable change as the starting point for research and assessment may not only offer a logical starting point for assessing success, but it may also offer a good opportunity to connect with policymakers’ thoughts and ambitions (whether publicly expressed or not).

The rise of sustainability means that both policymakers and evaluators must face up to more complex causal relationships between policy and benefits and, hence, use the alternative methods described to establish an evidence base between complex sets of interventions and outcomes. In addition, they must establish a ‘higher level’ orientation towards relevance and learning.

My conclusion is: Theories of change and theory-based evaluation are here to stay. They can be improved by including system change theories or theories of sustainable change. I see four future avenues for a world in which sustainable impact is what is wanted:

1. When we succeed in working towards theories of sustainable change and ‘system change theory-based evaluation’, the quest for sustainable development may in fact be a new lifeline for a useful employment of theory-based evaluation.

In complex fields like development cooperation, road safety, or environmental protection, policymakers increasingly opt for a ‘systemic approach’ as an intervention strategy. Herein, various policy measures and actors must ‘work together’, while taking account of contextual developments. These ‘system approaches’ must be made part of theories of change and policy evaluation. At the same time, people still need focus to organize their thinking, knowledge management, debates, and decision making. With system change theories we do justice to complex policy programmes and acknowledge the contributions from various network partners.

2. There are useful context-driven, ‘back tracing’ methods available for evaluating system change or ‘transition’.

Policies that seek to achieve sustainable change—whether in development cooperation, innovation policy, or road safety—require a system-based approach that is, by definition, more complex than the implementation of individual measures. As policymakers are moving from ‘simple projects’ to ‘complex development interventions’, evaluators are faced with growing complexity. They must be able to employ these alternative methods to do justice to contextual factors that support (or undermine) the sustainability of benefits, whether these can be assessed quantitatively or qualitatively. These include but are not limited to stakeholder ownership and engagement, absorptive capacity, political will, and long-term resource availability.

3. Within complex policy interventions, the deployment of ‘traditional’, RCT-like evaluations of individual policy measures still has great value.

The evaluation of individual policy measures is still useful, but mainly as part of ‘bigger’, sustainability-oriented evaluations on a system level. In *Less Pretension, More Realism* (IOB, 2019), the main finding of the evaluation was that two important programmes achieved several tangible outputs; but in the longer term, their impact was limited. While concrete results were achieved (such as the establishment of peace committees or infrastructural works), outcomes were often restricted to individual cases or were local in scope. In only a few instances did outcomes ‘trickle up’ and positively influence broader socioeconomic development or reduced levels of conflict or insecurity.

4. When seeking sustainable development, policymakers should be aware of the risk of ‘overreaching’.

The ever-increasing complexity of social and environmental problems is reflected in policy programmes. As the IOB evaluation demonstrates, it is useful to base interventions on explicit and tested assumptions and a sincere appraisal of the sustainability of expected outcomes. The conclusion may well be that more realism is needed: be aware of the risk of ‘overreaching’. Given the many variables that determine long-term success on a system level, policymakers should beware of ‘optimism bias’.

References

- Decouttere, C., De Boeck, K., & Vandaele, N. (2021). Advancing sustainable development goals through immunization: A literature review. *Global Health*, 17(1), 95.
- Douthwaite, B., Mayne, J., & McDougall, C. (2017). Evaluating complex interventions: A theory driven realist-informed approach. *Evaluation*, 23(3), 294–311.
- IOB. (2019). *Less pretension, more realism*. Ministry of Foreign Affairs of the Netherlands, Policy and Operations Evaluation Department, The Hague.

- Mansuri, G., & Rao, V. (2012). *Localizing development—Does participation work?* The World Bank.
- OECD. (2017). *Debate the issues: Complexity and policy making*. OECD Insights, OECD Publishing. <http://dx.doi.org/10.1787/9789264271531-en>
- OECD. (2021). *Applying evaluation criteria thoughtfully*. OECD Publishing. <https://doi.org/10.1787/543e84ed-en>
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage Publications.
- Scriven, M. (1974). Maximizing the power of causal investigations: The modus operandi method. In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 68–84). McCutchan Publishing.
- USAid. (2022). *Evaluating complex programs: Complex interventions and the challenges of evaluating them*. Module 17. USAid Learning Lab.
- Van der Knaap, P. (2003). Theory-based evaluation and learning: Possibilities and challenges. *Evaluation, 10*(1), 16–34.
- Van der Knaap, P. (2011). Sense and complexity: Initiatives in responsive performance audits. *Evaluation, 17*(4), 351–363.
- White, H., Menon, R., & Waddington, H. (2018). *Community-driven development: Does it build social cohesion or infrastructure? A mixed-method evidence synthesis*. Working Paper No. 30. International Initiative for Impact Evaluation.
- Wilson-Grau, R., & Britt, H. (2013). *Outcome harvesting*. Ford Foundation Mena Office.