

Three approaches to evaluation

Evaluation evolution?

Politicians are calling for evaluations that measure the effects of development cooperation. However, good development cooperation focuses on long-term processes that cannot be measured in terms of cause and effect. Alternative approaches to evaluation are needed.

Development cooperation is one of the most evaluated areas of public policy. Over the past 30 years, several typologies of evaluators and their approaches have been developed. 📖 Basically there are three types of evaluation: evidence-oriented evaluation, which seeks hard evidence; realistic evaluation, which tests how and why outcomes of policy occur; and complexity evaluation, which focuses on the complexity of social issues and governance (see table on page 25).

The evidence-oriented approach is the most dominant in development cooperation evaluation, but it is not necessarily the most illuminating. The realistic approach, meanwhile, is mainly applied in the fields of justice, health and social services in European countries. According to some theorists, however, a shift is occurring away from evidence-oriented to realistic and complexity evaluation. The importance of the social and political contexts in which a policy is employed is increasingly recognized. This context is dynamic, complex and multilayered, involving many different agencies and networks.

Complexity evaluation is related to the more recent use of complexity theory in social science. This emerging approach may provide useful insights to help overcome serious flaws in current evaluation practice, particularly in developing countries.

Evidence-oriented evaluation

The evidence-oriented evaluation approach aims to find measurable changes that can be directly attributed to specific policies. The approach uses ‘experimental’ research

methods. It is evaluation by testing, just as the effect of a medical treatment is assessed in laboratories by administering it to some members of a test group and not to others. The idea is that by using a large sample group, you can determine the effects of a programme or project objectively.

An example of evidence-oriented evaluation is the approach of the Independent Evaluation Group (IEG) of the World Bank. The IEG employs an ‘objectives-based approach’ in which impact evaluation is the most important element. According to the IEG, programmes and projects can be seen as ‘having a results chain – from the intervention’s inputs, leading to its immediate output, and then to the outcome and final impacts’.

But evidence-oriented evaluation takes for granted policy objectives. It does not reconstruct policy theory as a set of underlying assumptions about the relationships between aims, means and results, let alone attempt to identify and test the different and sometimes conflicting policy theories within a ministry.

Realistic evaluation

Realistic evaluation assesses how policy is ‘received’ under certain social and cultural conditions and how it triggers a response from people. To explain this method, Ray Pawson of the University of Leeds and Nick Tilley of Nottingham Trent University use the example of a watch. If you look at a watch at different times of the day, you can conclude that the hands move, but you cannot understand the underlying mechanism that makes them move. Pawson and Tilley have serious doubts about presenting input, output, outcome and impact in terms of cause and effect. They also think ‘policy theories’ that show the relationships between aims, means and outcomes with arrows offer a distorted view of reality.

These theories remain at best simplified cause and effect diagrams and, like the IEG and other evidence-oriented

By **Otto Hospes**: associate professor of law and governance at Wageningen University, the Netherlands. From 2002 until 2007 he was seconded to the Policy and Operations Evaluation Department (IOB) of the Netherlands Ministry of Foreign Affairs.



Alberto Estévez ANP Photo

A human domino chain in which people 'died' after a fake missile was dropped in protest against the war in Iraq. Barcelona, March 2003

evaluation units, assume the existence of a results chain.

As a first step, Pawson and Tilley suggest replacing the arrows used in evidence-oriented evaluation to link input, output, outcome and impact with black boxes. Each black box represents an organization, a social interaction or a social force field in which input is converted into output, output into outcome and outcome into impact. But there is not necessarily a linear connection.

The second step is to come up with a number of theories about what happens in or because of the black box and what possibly determines the outcomes of a programme. According to Pawson and Tilley, programmes work (i.e.

they have successful 'outcomes') only in so far as they introduce suitable ideas and opportunities ('mechanisms') to groups in the appropriate social and cultural conditions ('contexts'). 🇺🇸

Evaluators have traditionally questioned how to determine whether an observed change can be attributed to a particular policy. From the perspective of black boxes, mechanisms and contexts, this question of attribution becomes an exploration of the dynamics that can exist in a programme, social interaction or other factors that influence societies.

Realistic evaluation typically addresses very concrete subjects, primarily in Europe and North America. One example would be installing closed circuit television cameras in car parks to prevent crime. For such projects, different mechanisms and contexts are identified as possible reasons for outcomes. Through logical thinking and surveys, theories are proved invalid, and the realistic evaluator ends up with two or three probable combinations of mechanism and contexts. Hypothesis testing is something realistic evaluators have in common with evaluators using experimental methods. In fact, some consider realistic evaluation as simply a sophisticated version of a natural, scientific and evidence-based approach to evaluation.

Evaluation units in the field of development cooperation have not embraced realistic evaluation. These units are paid by governments that want definitive reports on concrete results of their policies in developing countries, rather than an unpredictable exploration of context and mechanisms explaining outcomes.

The relationship between policymakers and realistic evaluators is also rather tense. For many policymakers it is logical to see – or at least organize – policymaking as a rational cycle of design, implementation, evaluation and redesign. For realistic evaluators that is not at all self-evident. Digging around in a black box means that a ministry or implementing organization can also be the subject of an evaluation. According to Pawson, ➤

Evidence-oriented evaluation	Realistic evaluation	Complexity evaluation
Measuring effects	Investigating black boxes	Exploring complexity
No use of policy theory	Use of policy or programme theories on what happens in black boxes	Starting point is that policymaking is dynamic and interactive
Programmes stand at the beginning of a results chain	Programmes are black boxes	Programmes are adaptive systems
Input -> output -> outcome -> impact	Mechanism + context = outcome	Outcomes are emerging and quite unpredictable
One-way and single cause-effect relationships	One-way and multiple cause-effect relationships	Two-way and multiple cause-effect relationships

policy-makers do not want to hear that ‘we can never know, with certainty, whether a programme is going to work’. 📖

Complexity evaluation

The complexity approach views policy as a dynamic system and tries to evaluate how policy-makers respond to complex problems. This approach is a reaction to realistic evaluation, which is ironically seen as lacking in realism. Drawing lessons from extensive evaluation of complex policy initiatives, Marian Barnes of the University of Brighton, UK, says, ‘There is a danger that in the realistic evaluation formulation (context + mechanism = outcome), context is conceptualized as external to the programme being evaluated, and that it is addressed solely as factors which facilitate or constrain the achievement of objectives’. The criticism is that realistic evaluators, and their evidence-oriented colleagues, pay insufficient attention to the dynamics of the context, mechanism and outcome relationship.

Complexity can be divided into three distinct dimensions. The first categorizes problems into types. There are ‘wicked problems’, ‘cross-cutting issues’ and ‘complex problems’. ‘Wicked problems’, according to Barnes, are ‘those social and economic problems that demand interventions from a range of agencies and bodies, and which often seem intractable’. Such problems typically occur at several levels simultaneously. There are many wicked problems and cross-cutting issues in the field of development. Poverty, for example, cannot be addressed without considering issues such as peace and security, trade, agriculture or the environment.

The second dimension refers to dynamic systems. When several agencies and bodies adopt a joint approach to a wicked problem at different levels, the situation in which each agency wants to bring about change continually evolves. Unexpected circumstances can arise. Projects and programmes are not static: ‘Once a programme is in operation, the relationships between links in the causal hierarchy are likely to be recursive rather than unidirectional’. This is known as ‘recursive causality’, a concept invented by Patricia Rogers of RMIT University in Australia. Taking into account these ‘shifting goal posts’ – the adjustment of aims and strategies – is important. After all, target groups and intended beneficiaries of policy make the programme. They give it meaning and push it in a certain direction. In short, a programme or project is a sort of adaptive, emerging system.

The third dimension concerns other styles or methods of making policy with the aim of addressing complex systems and managing programmes as adaptive systems. Many policy-makers are of course aware of wicked problems and call repeatedly for an integrated approach or harmonization of interventions. Likewise, they know that a ministry or programme cannot be run just by pressing a button. They are aware that ‘in diverse dynamic and complex areas of society activity, no single governing agency is able to realize legitimate and effective governing by itself’, according to Gerry Stoker of the University of Southampton, UK. This

awareness has led to calls for ‘interactive governance’ as a different style of policy-making, one in which policy is developed in continual dialogue with representatives of the private sector, civil society organizations and education and research institutes. Evaluating the effectiveness and legitimacy of ‘interactive governance’ is no simple matter.

The drivers and scientific roots of complexity evaluation vary. Among evaluators, according to Marion Barnes, ‘there is now widespread acceptance that exclusively experimental models are inappropriate for the evaluation of complex policy initiatives that seek multi-level change with individuals, families, communities and systems’. When evaluating complexity, realistic evaluation is seen, at best, as a starting point. It is not considered to be sufficient on its own. The idea is not so much to combine different methods of research but to bring together different approaches, based on adaptive systems, institutional dynamics and assigning meaning. In other words, evaluations of complexity must be rooted in various disciplines.

Thinking only in terms of cause and effect is becoming obsolete. It will never be possible to determine precisely to what extent certain results or changes in the way societies or people behave can be attributed to a particular policy or programme. It is even less possible to determine this if the development effort is coming from a foreign country. Politicians and policy-makers should reject the illusion of predictability and not insist on real proof. The notions of causality and control must be replaced by notions of complexity and adaptation to complexity. Logic and reasoning should be applied to learning and to the development of alternative, more reflexive, styles of policy-making. ■

The author wishes to thank Arie de Ruijter, dean and professor at the Faculty of Humanities, Tilburg University, and Anneke Slob, director, Macro and Sector Policy Division, ECORYS, the Netherlands.

-
- Barnes, M. *et al.* (2003) Evidence, understanding and complexity. *Evaluation*, 9(3): 265–284.
 - Barrientos, Clarke, A. (1999) *Evaluation Research*. Sage.
 - Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. Sage.
 - Pawson, R. (2003) Nothing as practical as a good theory. *Evaluation*, 9:471–490.
 - Rogers, P. (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1): 29–48.
 - Sanderson, I. (2002) Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80(1): 1–22.

📖 A longer version of this article, with references and notes, can be found at www.thebrokeronline.eu.