

Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions

PATRICIA J. ROGERS
RMIT University, Melbourne, Australia

This article proposes ways to use programme theory for evaluating aspects of programmes that are complicated or complex. It argues that there are useful distinctions to be drawn between aspects that are complicated and those that are complex, and provides examples of programme theory evaluations that have usefully represented and address both of these. While complexity has been defined in varied ways in previous discussions of evaluation theory and practice, this article draws on Glouberman and Zimmerman's conceptualization of the differences between what is complicated (multiple components) and what is complex (emergent). Complicated programme theory may be used to represent interventions with multiple components, multiple agencies, multiple simultaneous causal strands and/or multiple alternative causal strands. Complex programme theory may be used to represent recursive causality (with reinforcing loops), disproportionate relationships (where at critical levels, a small change can make a big difference – a 'tipping point') and emergent outcomes.

KEYWORDS: collaboration; complexity; performance measurement; programme theory; theory of change

Introduction

Life is not simple, but many of the logic models used in programme theory evaluation are. Is it a problem to represent reality as a simple causal model of boxes and arrows, or should the logic models we use address the complexity of life – and if so, how?

There are significant challenges for programme theory when it is used to evaluate interventions with complex aspects, and not all evaluators have been convinced it is feasible or useful. Stufflebeam (2004: 253), for example, has argued that programme theory evaluation makes little sense because it 'assumes that the complex of variables and interactions involved in running a project in the complicated,

sometimes chaotic conditions of the real world can be worked out and used a priori to determine the pertinent evaluation questions and variables'. Pinnegar (2006: 4), while welcoming more holistic and thoughtful responses to complex issues such as meeting housing needs, has gone further and questioned the value of complex programmes altogether, suggesting that a complex programme tends to be 'too difficult to explain its objectives in tangible terms, too amorphous to deliver, and too difficult to meaningfully evaluate'.

Other evaluators, however, have found ways to address the challenges of complicated and complex aspects of interventions – both through the types of logic models that are used and how they are used – and this article brings together a number of published evaluations that can be used as examples. Importantly, these examples do not involve creating messier logic models with everything connected to everything. Indeed, the art of dealing with the complicated and complex real world lies in knowing when to simplify and when, and how, to complicate.

Programme Theory and Complexity Theory

Programme theory, variously referred to as programme theory, programme logic (Funnell, 1997), theory-based evaluation or theory of change (Weiss, 1995, 1998), theory-driven evaluation (Chen, 1990), theory-of-action (Schorr, 1997), intervention logic (Nagarajan and Vanheukelen, 1997), impact pathway analysis (Douthwaite et al., 2003b), and programme theory-driven evaluation science (Donaldson, 2005) refers to a variety of ways of developing a causal modal linking programme inputs and activities to a chain of intended or observed outcomes, and then using this model to guide the evaluation (Rogers et al., 2000). In this article, the term 'logic model' is used to refer to the summarized theory of how the intervention works (usually in diagrammatic form) and 'programme theory evaluation' is used for the process of developing a logic model and using this in some way in an evaluation.

Most approaches to building logic models have focused on simple, linear models, but some have explored how non-linear models might be used (e.g. Funnell, 2000; Rogers, 2000) to better represent programmes and guide their evaluation. In particular, a number of evaluators have incorporated concepts of complexity in their discussion and use of logic models (e.g. Barnes et al., 2003; Davies, 2004, 2005; Douthwaite et al., 2003a, 2003b; Pawson, 2006; Sanderson, 2000; Stame, 2004) but have defined the terms in quite different ways and addressed different aspects of complexity and of programme theory. This article presents a framework for classifying the different aspects of complexity that might be addressed in programme theory and examples of how this might be done.

While there have been many diverse definitions and conceptualizations of complexity (e.g. Eoyang and Berkas, 1998; Kurtz and Snowden, 2003; Stacey, 1996), the starting point for this article is the three-part distinction (Glouberman, 2001; Glouberman and Zimmerman, 2002) between what is complicated (lots of parts) and what is complex (uncertain and emergent). Their distinction is summed up in the widely used comparative table shown in Table 1.

Table 1. Simple, Complicated and Complex Problems (Glouberman and Zimmerman, 2002)

<i>Simple:</i> <i>Following a recipe</i>	<i>Complicated:</i> <i>Sending a rocket to the moon</i>	<i>Complex:</i> <i>Raising a child</i>
The recipe is essential	Formulae are critical and necessary	Formulae have a limited application
Recipes are tested to assure easy replication	Sending one rocket to the moon increases assurance that the next will be OK	Raising one child provides experience but no assurance of success with the next
No particular expertise is required but cooking expertise increases success rate	High levels of expertise in a variety of fields are necessary for success	Expertise can contribute but is neither necessary nor sufficient to assure success
Recipes produce standardized products	Rockets are similar in critical ways	Every child is unique and must be understood as an individual
The best recipes give good results every time	There is a high degree of certainty of outcome	Uncertainty of outcome remains
Optimistic approach to problem-solving	Optimistic approach to problem-solving	Optimistic approach to problem-solving

This distinction seemed to provide a particularly useful orienting framework for evaluation (Patton, 2003) and has obvious implications for programme theory evaluation. The evaluation of simple interventions is intended to either develop or test the ‘recipe’ that others can then follow. Complicated interventions that have many components pose challenges to evaluations, given the limited number of variables that can be identified and empirically investigated. But it is complex interventions that present the greatest challenge for evaluation and for the utilization of evaluation, because the path to success is so variable and it cannot be articulated in advance.

While the Glouberman and Zimmerman classification refers to types of problems, and some applications of their work have used it to classify interventions or service systems (e.g. Martin and Sturmberg, 2005, classify general medical practices as ‘locally run complex systems’), it is probably more useful to consider these classifications as different ways of looking at interventions, and to classify aspects of them rather than the interventions themselves. A complex intervention may well have some simple aspects where known causal processes are put in train.

Classification of Issues

Table 2 sets out the five issues that are discussed in this article, their implications for evaluation, and a suggested classification as either complicated or complex aspects of interventions.

Table 2. Complicated and Complex Aspects of Interventions

Aspect	Simple version	Not-simple version	Challenges for evaluation	Suggested label
1. Governance and implementation	Single organization	Multiple agencies, often interdisciplinary and cross-jurisdictional	More work required to negotiate agreement about evaluation parameters and to achieve effective data and collection analysis	Complicated
2. Simultaneous causal strands	Single causal strand	Multiple simultaneous causal strands	Effective programs may need to optimize several causal paths, not just one; evaluation should both document and support this	Complicated
3. Alternative causal strands	Universal mechanism	Different causal mechanisms operating in different contexts	Replication of an effective programme may depend on understanding the context that supports it. The counter-factual argument may be inappropriate when there are alternative ways to achieve the outcome	Complicated
4. Non-linearity and disproportionate outcomes	Linear causality, proportional impact	Recursive, with feedback loops	A small initial effect may lead to a large ultimate effect through a reinforcing loop or critical tipping point	Complex
5. Emergent outcomes	Pre-identified outcomes	Emergent outcomes	Specific measures may not be able to be developed in advance, making pre- and post-comparisons difficult	Complex

The remainder of the article presents and discusses examples of programme theory evaluations that address each of these issues through the type of logic model developed and how it was used for the evaluation.

Methodology

A three-part search was undertaken to explore how these issues had been addressed in programme theory evaluations. A textual search was made of published literature using the search terms 'program theory' (and its many synonyms) and 'complexity or complex'. A visual search was made of images on the internet using Google and the previous search terms. Other examples obtained or developed in the course of the author's evaluation projects were also reviewed in terms of the issues raised by the framework. The examples presented in this article are intended to be illustrative rather than presented as best practice, and to encourage others to share other examples which address these issues more effectively.

Simple Logic Models

Many logic models used in programme theory (and guides to developing programme theory) show a single, linear causal path, often involving some variation on five categories (inputs, processes, outputs, outcomes and impact). Figure 1 shows a particularly influential version of this in a guide to developing and using logic models published by the W. K. Kellogg Foundation.

For some interventions, simple logic models such as these are quite appropriate. There is, however, currently considerable debate about how appropriate this is for many human service interventions such as education, drug prevention, family support services and international development.

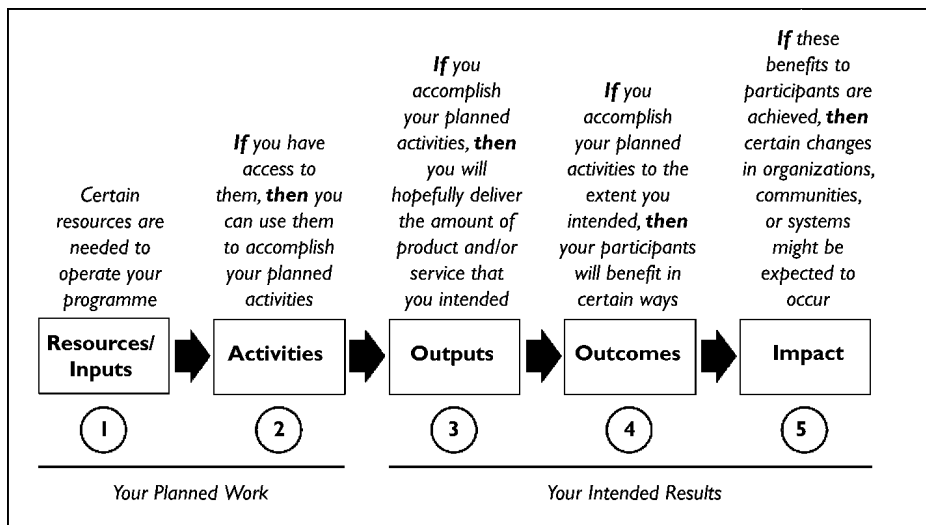


Figure 1. A Simple Logic Model (W. K. Kellogg Foundation, 2004)

While these models provide a clear statement of the overall intent of an intervention, and useful guidance for implementation and selection of variables for an evaluation, there can be risks in using them, particularly if they are taken literally when they are not exactly true. By leaving out the other factors that contribute to observed outcomes, including the implementation context, concurrent programmes and the characteristics of clients, simple logic models risk overstating the causal contribution of the intervention, and providing less useful information for replication. Where they are used to develop performance targets for accountability, there is the risk of 'goal displacement' (Perrin, 2003; Winston, 1999), where original targets are met even though this undercuts the actual goals of the intervention.

Barnes et al. (2004: 13–14) have pointed out that the use of models such as these for evaluation assumes 'a stable environment in which any indication of either theory or implementation failure would be capable of adjustment in line with available evidence'.

Simple logic models are also more likely to present a single theory of change, rather than representing different stakeholders' views about what are desirable outcomes and how these might be achieved. For example, Bakewell and Garbutt (2005: 19), in their review of the logical framework approach, have pointed to the way in which it encapsulates a particular type of theory of change:

... the LFA encapsulates a theory of change – a linear model of inputs causing a set of predictable outcomes. Some might go further and say that it is being used as a tool to impose a set of development ideas on communities in developing countries. As such it represents an ideology rather than being an objective, technical management tool.

Eoyang et al. (1998: 3) have warned of the dysfunctional effects when people try to use a simple, linear model for planning and evaluating an intervention that is more like a complex adaptive system:

Everyone involved in making public policy can think about the process as if it were well regulated and linear. Their project plans and shared discourse may revolve around the orderly steps of the problem solving method, which is their espoused theory. In reality, however, they experience the process as a surprising, uncontrolled, and emergent phenomenon. This distinction between espoused theory and experience leads to a variety of unpleasant outcomes. Participants blame themselves or others when the process does not progress as it should. Resources are wasted in pursuit of the perfect and controlled response. Opportunities are missed when serendipity is damped or ignored because it does not fit in the expected scheme. Personal and professional frustration result when well laid plans prove ineffective.

Given these issues, Wholey (2003: 10) has outlined a relatively narrow range of interventions that are suitable for results-oriented management that is based on a simple logic model – those where goals can be agreed and precisely quantified, where progress towards them can be reliably measured, and where both staff activities and the results of those activities can be readily observed (production agencies such as the Postal Service). In light of this observation, it is instructive to notice how many guides to using logic models to develop performance indicators use production agencies as their examples, and do not discuss the challenges in

translating these examples to programmes that do not directly deliver services, or where the activities or the results of these activities cannot be readily observed.

Gregory (1995: 58) has gone further and argues that:

Treating all tasks as if they were amenable to a production culture not only is likely to have counter-productive effects with regard to goal displacement, but may also encourage official behavior which, while accountable, is less responsible, even corrupt.

These simple models might therefore best be reserved either for aspects of interventions that are in fact tightly controlled, well-understood and homogeneous or for situations where only an overall orientation about the causal intent of the intervention is required, and they are clearly understood to be heuristic simplifications not accurate models.

Complicated Logic Models

Aspects of Complication

Three aspects of complication have been addressed in published examples of evaluations: interventions implemented through multiple agencies; interventions with multiple simultaneous causal strands; and interventions with alternative causal strands (see Table 3). Each of these will be discussed.

Multi-Site, Multi-Governance

One way in which interventions can be complicated, without necessarily being complex, is when they are implemented in different sites and/or under different governances. For example, the World Bank's (2004) account of efforts since 1974 to control the parasitic worm that causes onchocerciasis, or river blindness, in Africa refers to the partnership of eleven national governments, four international sponsors, and a host of private-sector companies and non-government organizations. Clearly, this presents challenges in terms of reaching agreement about evaluation planning, methods for data collection and analysis, and reporting to different audiences. However, there was a good understanding of the causal path for the intervention. Scientists knew which parasitic worm caused the problem and how its lifecycle could be interrupted. Although there was some local adaptation of implementation, there was an overall plan that was understood and agreed.

Table 3. Three Aspects of Complication

<i>Aspect of complication</i>	<i>Simple intervention</i>	<i>Complicated intervention</i>
1. Governance and location	Single organization	Multiple agencies, often interdisciplinary and cross-jurisdictional
2. Simultaneous causal strands	Single causal strand	Multiple simultaneous causal strands
3. Alternative causal strands	Universal mechanism	Different causal mechanisms operating in different contexts

The challenges for evaluation of complicated interventions with multiple governance are primarily logistical. This is not to understate the scale of work involved in such interventions, or in evaluations of them. In terms of logic models, however, a single logic model might be used, with data reported separately for each jurisdiction, and then aggregated for an evaluation of the overall intervention. This is not true of other types of complicated interventions or complex interventions.

Simultaneous Causal Strands

A second aspect of complication is the existence of two or more simultaneous causal strands that are all required in order for the intervention to succeed. It can be important for a logic model to show both of these, and for the evaluation to gather data about both of them, so that important parts of the intervention are not left out.

Simultaneous causal strands were an important aspect addressed in the monitoring and evaluation of a maternal and child health service (Rogers, 2004; as shown in Figure 2). The programme sought to support parents to develop confidence in parenting at the same time as trying to encourage them to adopt healthier nutritional practices. Programme staff believed it was important to show in the logic model the need for balance between these two causal paths, to explain why they could not simply focus on strong messages about nutrition.

It was therefore seen as necessary to show two different causal strands that were sometimes in tension – a strand where the expert staff provided skills and knowledge to the parents; and a strand where staff supported parents to have confidence in their parenting abilities. This was an important point when using the

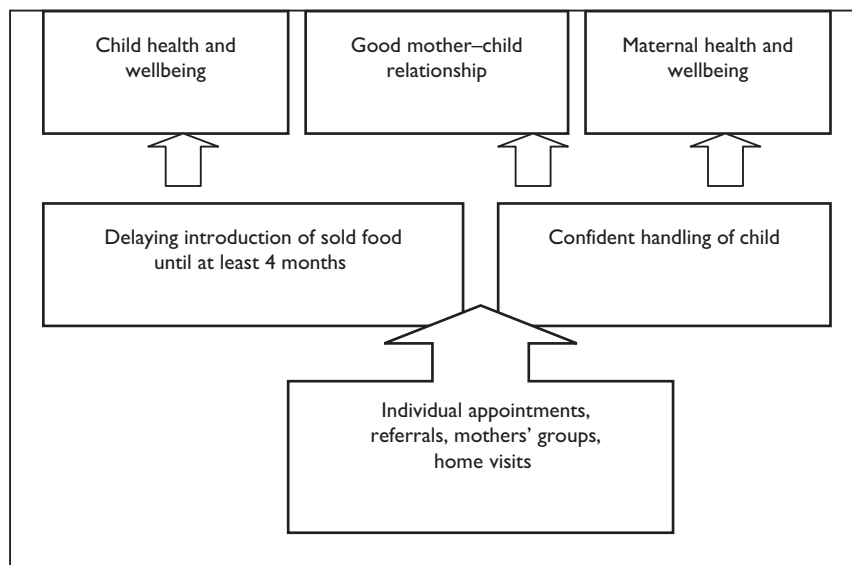


Figure 2. Logic Model Showing Simultaneous Causal Strands

logic model as a communication device to explain the intervention to new staff or to other agencies, and when deciding which aspects should be included in monitoring or evaluation. To make it clear that these causal strands are not optional alternatives but each essential, it might be better to represent them using arrows that show clearly that both are required, perhaps as in Figure 2.

An intervention with multiple simultaneous causal strands cannot afford to only focus on achieving one of these. An evaluation needs to both document and support this.

The previous example of controlling river blindness has something of this element as well. The success of the intervention depended on achieving success in each of the different locations (otherwise worms would simply breed in some areas and reinfest the others). It may have therefore been important to show simultaneous causal strands for each of the sites, all of which were necessary to achieve the final outcome.

Alternative Causal Strands

A third aspect of interventions that can be considered complicated involves alternative causal strands. In most logic models, these appear as parallel lines of boxes and arrows and are visually indistinguishable from simultaneous causal strands. The difference is that a programme can work either through one or other of the causal paths. In many cases, these alternative causal strands are effective in particular contexts – the ‘what works for whom in what ways’ notion of realist evaluation (Pawson and Tilley, 1997). So closed circuit television may work to reduce automobile theft in different ways in different contexts – through passive surveillance where it increases usage of car parks with spare capacity, or through capturing and removing thieves in contexts where it is part of an enforcement programme and where there is not a large pool of potential thieves to fill the void once some are removed.

Sanderson (2000: 447) has focused on the issue of context dependency, in terms of evaluating complex policy systems:

... [C]omplexity theory suggests that the effect of a policy intervention may be due to the fact that it was ‘fortuitously’ introduced at a particular point in time into a particular set of circumstances such that it provided the catalyst for a step change in the system.

In addition to leading to a need for logic models that show different causal paths in different contexts, or a causal path that is only effective in favourable contexts, this aspect of complexity leads to Sanderson’s suggestion to use these to undertake evaluations that involve ‘comparative analyses over time of carefully selected instances of similar policy initiatives implemented in different contextual circumstances’.

Glouberman and Zimmerman’s archetype for a complicated intervention, the rocket ship, might be understood as being complicated only in terms of the number of components to coordinate, but it is possible to consider it also as complicated in terms of the differential prescriptions for practice depending on the context. For the rocket ship, environmental conditions such as temperature affect the causal path involved in launching and hence should be addressed in monitoring and evaluation.

For general medical practice, other existing medical conditions are an important complicating factor that affects the causal path involved in taking certain drugs and hence should be addressed (Martin and Sturmberg, 2005).

Alternative causal strands can also be important to document in evaluation to guide appropriate replication into other locations and times. These can be difficult, however, to show diagrammatically and few examples exist that show these clearly. Instead this type of logic model has more commonly been represented in a tabular form.

Complex Interventions and Logic Models

Aspects of Complexity

Two aspects of complexity have been addressed in published examples of evaluations: recursive causality and emergence (see Table 4). These will now be discussed.

Recursive Causality and Tipping Points

Patton (1997: 232) points out that, although logic models usually show a linear progression from initial outcomes to subsequent outcomes,

Once a program is in operation, the relationships between links in the causal hierarchy are likely to be recursive rather than unidirectional. The implementation and attainment of higher-level objectives interact with the implementation of lower-order objectives through feedback mechanisms [and] interactive configurations. . . . In short, the cause-effect relationships may be mutual, multidirectional and multilateral.

Most logic models show ‘one pass’ through the intervention, but many interventions depend on activating a ‘virtuous circle’ where an initial success creates the conditions for further success. This means that evaluation needs to get early evidence of these small changes, and track changes throughout implementation.

‘Tipping points’, where a small additional effort can have a disproportionately large effect, can be created through virtuous circles, or be a result of achieving certain critical levels. For example, Batterham et al. (1996), after reviewing studies of recovery from various disabling conditions, found that certain outcome states could be considered thresholds that make the outcome usable and, hence, sustainable or that create the opportunity for further improvement. The causal path shown in a logic model might only occur at critical levels of activity, or once certain thresholds of earlier outcomes have been achieved. This can be difficult to

Table 4. Two Aspects of Complication

<i>Aspect of complexity</i>	<i>Simple intervention</i>	<i>Complex intervention</i>
1. Recursive causality and disproportionate effect	Linear, constant dose-response relationship	Recursive, with feedback loops, including reinforcing loops; disproportionate effects at critical limits
2. Emergent outcomes	Pre-identified outcomes	Emergent outcomes

show diagrammatically, and is perhaps best communicated through annotations on a diagram.

Emergence

One of the most challenging aspects of complex interventions for evaluation is the notion of emergence – not that certain patterns emerge as our understanding of them improves (knowledge which can then be used to predict similar interventions in the future), but that the specific outcomes, and the means to achieve them, emerge during implementation of an intervention.

Regine and Lewin (n.d.) describe complex adaptive systems as follows:

Simply defined, complex adaptive systems are composed of a diversity of agents that interact with each other, mutually affect each other, and in so doing generate novel behavior for the system as a whole. But the pattern of behavior we see in these systems is not constant, because when a system's environment changes, so does the behavior of its agents, and, as a result, so does the behavior of the system as a whole. In other words, the system is constantly adapting to the conditions around it. Over time, the system evolves through ceaseless adaptation.

Rather than being a symptom of ineffective management, emergent outcomes might be appropriate in the following cases:

- when dealing with a 'wicked problem' – intractable, don't know how to deal with it;
- where partnerships and network governance are involved, so activities and specific objectives emerge through negotiation and through developing and using opportunities (Uusikylä and Valovirta, 2004);
- where the focus is on building community capacity, leadership, etc., which can then be used for various specific purposes.

Emergent outcomes may well require an emergent logic model – or in fact one that is expected to continue to evolve. Patton (1994: 313) has called this 'developmental evaluation', arguing that:

Developmental programming calls for developmental evaluation in which the evaluator becomes part of a design team, helping to monitor what's happening, both processes and outcomes, in an evolving, rapidly changing environment of constant feedback and change.

This is quite different to some versions of programme theory. Indeed, the antecedent approach to programme theory, evaluability assessment, was based on the premise that, without an explicit model laying out the goals and specific measurable objectives, and how they are linked to programme activities, a programme could not be evaluated (Smith, 1989).

For emergent interventions, such as comprehensive community initiatives (Weiss, 1995), a series of logic models can be developed alongside development of the intervention, reflecting changes in the understanding. Data collection, then, must be similarly flexible. While the overall goals may be clear (e.g. building stronger families and communities), the specific activities and causal paths are expected to

evolve during implementation, to take advantage of emerging opportunities and to learn from difficulties. For these types of projects, a more flexible theory of change evaluation is needed, where an initial model is developed, and then both used to guide planning and implementation, but also revised as plans change. It can be difficult, however, for this local, flexible use of theory of change to meet the evaluation needs of a large, multi-site programme, where some common framework is needed across all projects (horizontal synthesis) and at different levels of outcomes (vertical synthesis) – that is, when interventions have aspects that are both complicated and complex. This combination of complication and complexity is examined next.

Interventions that have Both Complicated and Complex Aspects

The greatest challenge comes when interventions have both complicated aspects (multi-level and multi-site) and complex aspects (emergent outcomes). This is when a logic model needs to provide a common framework that can accommodate local adaptation and change.

Barnes et al. (2003, 2004) have discussed the difficulties they had in applying a theory of change approach to the evaluation of Health Action Zones (HAZ), despite early optimism about it (Judge and Bauld, 2001). HAZs involved at least seven dimensions of complexity: structural (with horizontal and vertical partnerships); temporal (with an aim of long-term changes but a need for short-term results); scope (intended catalytic organizational change being difficult to identify as within the scope of the intervention); multiple stakeholders with different perspectives; different theories of change evident across projects and also within projects; limitations in terms of required procedures; and a problem area (health inequalities) with multiple and contested causes.

They concluded:

The implications of such complexity are at the very least that multiple theories need to be articulated in respect of the multiple processes and relationships involved in delivering change. However our experience of evaluating HAZ leads us to suggest that this evaluation stretches the application of ‘Theories of Change’ to a point at which it becomes both methodologically and theoretically fragile. (Barnes et al., 2004: 13)

Similarly Kankare (2004), in his review of the evaluation of the European Social Fund (ESF) in Finland during 1995–9, concluded that it had increased complexity and ambiguity and reduced transparency, with the effect that the evaluations focused on the achievement of outputs, such as innovation and the development of networks to the exclusion of their connection to reducing unemployment.

Some evaluators have described more successful efforts to use multiple and emerging theories of change in programmes with multiple levels of stakeholders and emergent programmes.

Douthwaite et al. (2003a: 53), working in agricultural development in Africa, described innovation as a complex process characterized by ‘interactions among agents and processes, strong feedback loops and intrinsic randomness’. The logic model developed for the evaluation of an integrated striga management programme (Douthwaite and Schulz, 2001) addressed the combination of challenges. Striga is

a parasitic weed that is the severest biological constraint to cereal production in sub-Saharan Africa. Unfortunately, while there is an emerging body of knowledge about striga, its effective management requires an integrated approach, and knowledge about this is still being developed. At the same time, the programme is being implemented at multiple sites, and it is clear that local conditions affect what constitutes an effective integrated approach. The logic model in Figure 3 shows an overall theory of change with iterations of building and using knowledge, collaboratively between villages and researchers.

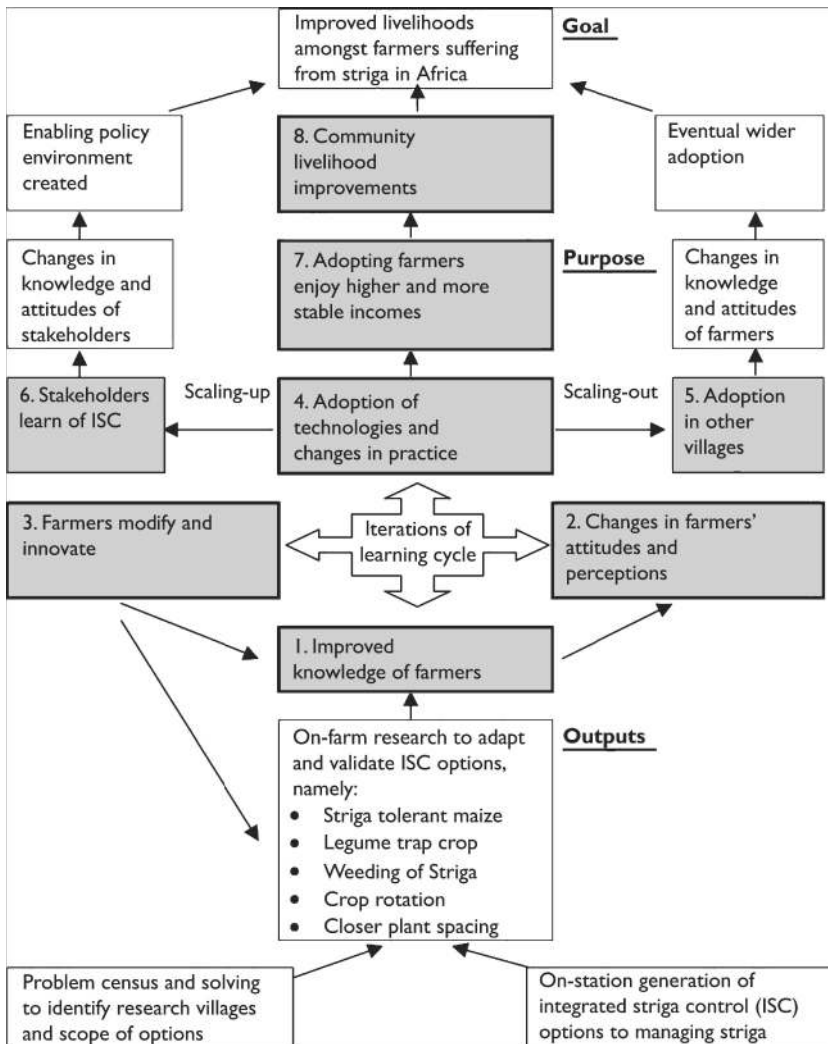


Figure 3. Logic Model for a Complicated and Complex Intervention: Striga Management (Douthwaite et al., 2003b)

Stame's (2004) discussion of the 'vertical complexity' of multi-level governance, such as in European Union programmes which involve local or subnational bodies implementing projects, relates to both the issue of multiple governance and the issue of alternative causal strands. She points out that evaluations implemented at the local level do not immediately inform evaluations of the impact of the overall programme. Weiss's (1998) distinction between implementation theory and causal theory can be useful here. The overall programme may have a common causal theory; particular sites (or particular contexts) may have different implementation theories of how to activate this causal theory.

Riggan (2005) developed what he termed a 'complex program theory' to guide the evaluation of community schools, a complex intervention involving a range of partnerships of organizations working together to achieve improved student learning and stronger families through services, supports and opportunities. To address the diversity and emergence of this intervention, the logic model was sufficiently broad to encompass the various stakeholders' different and emerging theories of change and focused on the collaboration as a fundamental component of the programme.

In a similar vein, a logic model was developed for use both in individual community capacity-building projects and across a national funding programme for such projects (CIRCLE, 2006). The diagram was generic enough to be relevant for a wide range of projects (including community leadership development, family strengthening, volunteer training and community events), and able to incorporate specific emergent outcomes as projects acted to capitalize on opportunities and address new issues.

Another option that has been recommended for this situation is to explore the use of network theory (Davies, 2004, 2005) to represent the heterarchical relationships between organizations, and between projects and overall programme goals.

Finally, a quite different approach is not to present a causal model at all, but to articulate the common principles or rules that will be used to guide emergent and

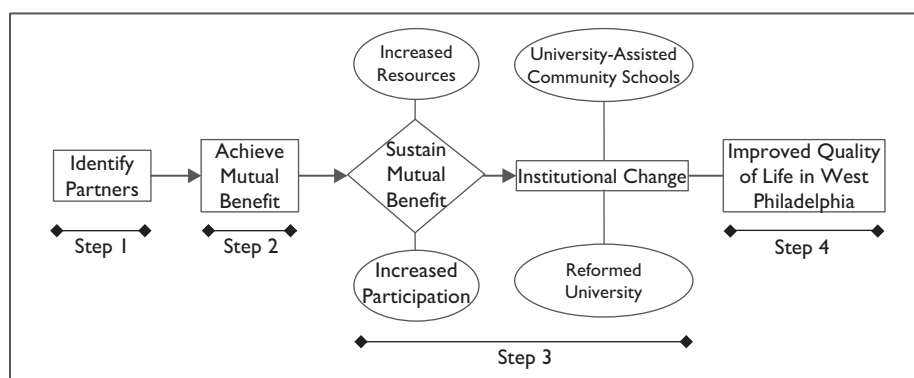


Figure 4. Logic Model for an Intervention with Complex and Complicated Aspects (Riggan, 2005)

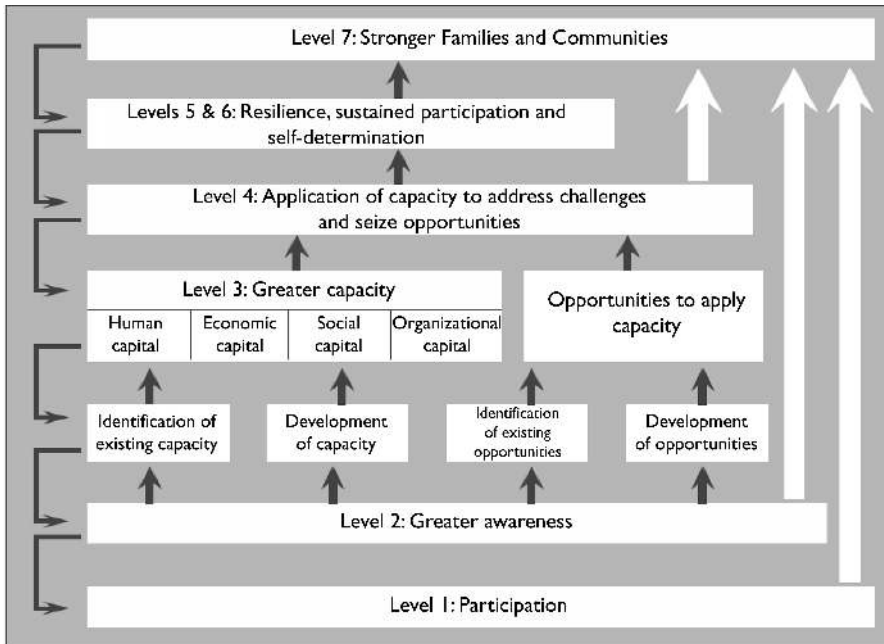


Figure 5. Logic Model for a Complex, Complicated Intervention: Community Capacity-Building Programme (CIRCLE, 2006)

responsive strategy and action. Eoyang and Berkas (1998) discuss this as one of the characteristics of Complex Adaptive Systems:

A complex adaptive system (CAS) consists of interdependent agents. The behavior of each agent conforms to a short list of simple rules, and the group of agents exhibits emergent, system-wide patterns of behaviour.

For example, the National Staff Development Council (Sparks, 2003) described its theory of change in terms of seven assumptions about the world, such as:

- Leaders matter. Reform that produces quality teaching in all classrooms requires skillful leadership at the system and school levels. Such leadership is distributed broadly to teachers and others in the school community.
- Significant change in teaching and learning begins with significant change in what leaders think, say, and do. The goal of leadership development, therefore, is deep understanding, transformational learning at the level of beliefs, and an unrelenting action orientation in the application of this understanding within a context of interpersonal accountability.

This is an example of the approach discussed by Sibthorpe et al. (2004: 3) with reference to healthcare systems.

When a new system is being instituted, a short list of simple rules (or minimum specifications) may be the most effective way to bring about change. They set the

parameters and provide both focus and freedom for system activities. Over-prescription is counter-productive because it stifles creativity and innovation.

Using Programme Theory Appropriately for Simple, Complicated and Complex Program Models

As well as developing appropriate logic models, we need to think carefully about how they are used. Particular care should be taken to not imagine that a logic model, however detailed, can be used to generate performance measures that can be used formulaically to modify implementation and improve performance when interventions have complex aspects.

Evaluators working on interventions with complicated and complex aspects have emphasized the importance of the discussions based around the logic models. For example, Potter (2004), using the example of an evaluation conducted by a network of practitioners and researchers engaged in urban regeneration in eight European cities, argues that evaluative methodologies for cluster evaluation of heterogeneous programmes, where results cannot simply be aggregated, need to be qualitative, communicative, iterative and participative. The method moved between field visits and case studies of individual cities to thematic analyses and final recommendations, with review and reframing by the group throughout.

In similar vein, Arnkil et al. (2002) advocate the use of 'emergent evaluation' to respond to 'glocalisation' (where decision-making is both pulled upwards to transnational networks and downwards to regional and local networks) and fuzzier objects and purposes of evaluation. For their evaluation, they developed a system called 'Good Future Dialogue' where people were asked to 'remember the future' – to imagine the future if the project had been successful, to describe this future, the support they had received to achieve this success, and the worries they had had along the way and how they been addressed. In some ways this sounds similar to some of the techniques used to develop logic models. Their process was both highly participative and recognized difference instead of seeking consensus that might reflect power differences rather than agreement. These multi-stakeholder dialogues were used simultaneously in the roles of data collection, hypothesis testing and intervention, rather than evaluators going away with the model and returning at the end with results. The future perspective is used to create an optimistic framework to tackle difficult and complex problems, and to suggest what they call a 'solution-oriented network', outlining specific ways forward, including leveraging of opportunities. This can be seen as a way to both develop and use emergent programme theory, as the stakeholders are together in real time during this process, and can start to use the emerging programme theories (as there is no intent to achieve consensus on just one) to guide planning, management and evaluation of their specific activities.

In the light of these issues, it is interesting to read Giuliani's (2002: 69–94) account of the use of the COMPSTAT data system in the New York Police Department, which shows the importance of the daily meetings to review and discuss the figures, as they worked iteratively on ways to improve performance.

The programme theory evaluation of the striga programme, previously discussed, also included participatory monitoring and evaluation to build better understanding and better implementation of the intervention. Douthwaite et al. (2003b: 262) concluded: 'Self-evaluation, and the learning it engenders, is necessary for successful project management in complex environments.'

Conclusion

Abma and Noordegraaf (2003) have discussed the challenges that uncertainty and ambiguity present for public sector managers. The anxiety provoked by uncertainty and ambiguity can lead managers and evaluators to seek the reassurance of a simple logic model, even when this is not appropriate. This article suggests that a better way to contain this anxiety might be to identify instead the particular elements of complication or complexity that need to be addressed, and to address them in ways that are useful, using the examples in this article as an initial guide.

Note

This is a substantially revised and expanded version of a presentation to the 2004 conference of the European Evaluation Society in Berlin. Thank you to referees, the editor and colleagues for their constructive comments on earlier versions.

References

- Abma, T. A. and M. Noordegraaf (2003) 'Public Managers Amidst Ambiguity: Towards a Typology of Evaluative Practices in Public Management', *Evaluation* 9(3): 285–306.
- Arnkil, R., T. Spangar, J. Nieminen, P. Rissanen and J. Konkinen (2002) *Emergent Evaluation and Learning in Multi-Stakeholder Situations*. Paper presented at the 2002 Conference of the European Evaluation Society, Seville. Available at <http://www.europeanevaluation.org/docs/ArnkilSevilleslides13.pdf>
- Bakewell, O. and A. Garbutt (2005) *The Use and Abuse of the Logical Framework Approach*. Stockholm: SIDA.
- Barnes, M., E. Matka and H. Sullivan (2003) 'Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems', *Evaluation* 9(3): 265–84.
- Barnes, M., H. Sullivan and E. Matke (2004) *The Development of Collaborative Capacity in Health Action Zones: A Final Report from the National Evaluation*. Birmingham: University of Birmingham. URL: http://www.socialresearch.bham.ac.uk/downloads/HAZ_FINAL_REPORT_JAN_11.pdf (downloaded 11 October 2007).
- Batterham, R., D. Dunt and P. Disler (1996) 'Can we Achieve Accountability for Long-Term Outcomes in Rehabilitation?', *Archives of Physical Medicine Rehabilitation* 77(12): 1219–25.
- Chen, H. T. (1990) *Theory-driven evaluation*. Thousand Oaks, CA: Sage.
- Collaborative Institute for Research, Consulting and Learning in Evaluation (CIRCLE) (2006) *Evaluation of the Stronger Families and Communities Strategy 2000–2004*. Third newsletter. Available at www.facsia.gov.au

Evaluation 14(1)

- Davies, R. (2004) 'Scale, Complexity and the Representation of Theories of Change (Part I)', *Evaluation* 10: 101–21.
- Davies, R. (2005) 'Scale, Complexity and the Representation of Theories of Change (Part II)', *Evaluation* 11: 133–49.
- Donaldson, S. (2005) 'Using Program Theory-Driven Evaluation Science to Crack the Da Vinci Code', in M. C. Alkin and C. A. Christie (eds) *Theorists' Models in Action*, New Directions in Evaluation, 106, pp. 65–84.
- Douthwaite, B., R. Delve, J. Ekboir and S. Twomlow (2003a) 'Contending with Complexity: The Role of Evaluation in Implementing Sustainable Natural Resource Management', *International Journal of Agricultural Sustainability* 1(1): 51–66.
- Douthwaite, B., T. Kuby, E. van de Fliert and S. Schulz (2003b) 'Impact Pathway Evaluation: An Approach for Achieving and Attributing Impact in Complex Systems', *Agricultural Systems* 78: 243–65.
- Douthwaite, B. and S. Schulz (2001) 'Spanning the Attribution Gap: The Use of Program Theory to Link Project Outcomes to Ultimate Goals in Integrated Natural Resource Management and Integrated Pest Management', paper presented at the INRM Workshop, 2001, Cali, Colombia. Available at <http://www.ciat.cgiar.org/inrm/workshop2001/docs/titles/4-1APaperBDouthwaite.pdf>.
- Eoyang, G. and T. Berkas (1998) 'Evaluation in Complex Adaptive System', in M. Lissack and H. Gunz (eds) *Managing Complexity in Organizations*. Westport, CT: Quorum Books.
- Eoyang, G., L. Yellowthunder and V. Ward (1998) 'A Complex Adaptive Systems (CAS) Approach to Public Policy Decision Making', paper presented to the Society for Chaos Theory in Psychology in the Life Sciences, August.
- Funnell, S. (1997) 'Program Logic: An Adaptable Tool for Designing and Evaluating Programs', *Evaluation News and Comment* 6(1): 5–7.
- Funnell, S. (2000) 'Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring', in P. Rogers, T. Hacsí, A. Petrosino and T. Huebner (eds) *Program Theory in Evaluation: Challenges and Opportunities*, New Directions for Evaluation, 87, pp. 91–101. San Francisco, CA: Jossey-Bass.
- Giuliani, R. (2002) *Leadership*. London: Little, Brown.
- Glouberman, S. (2001) 'Towards a New Perspective on Health Policy', CPRN Study No. H/03, Canadian Policy Research Networks Inc., Ottawa.
- Glouberman, S. and B. Zimmerman (2002) *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Commission on the Future of Health Care in Canada*, Discussion Paper 8. Available at http://www.hc-sc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman_E.pdf
- Gregory, R. J. (1995) 'Accountability, Responsibility, and Corruption: Managing the "Public Production Process"', in J. Boston (ed.) *The State Under Contract*, pp. 56–77. Wellington, NZ: Bridget Williams Books.
- Judge, K. and L. Bauld (2001) 'Strong Theory, Flexible Methods: Evaluating Complex Community-Based Initiatives', *Critical Public Health* 11(1): 19–38.
- Kankare, I. (2004) 'Treading on Fuzzy Grounds: Evaluation Practices in Managing Complexities', paper presented at the European Evaluation Society conference, Berlin.
- Kurtz, C. F. and D. J. Snowden (2003) 'The New Dynamics of Strategy: Sense-making in a Complex and Complicated World', *IBM Systems Journal* 42(3): 462–83. URL: <http://www.research.ibm.com/journal/sj/423/kurtz.pdf> (accessed 11 October 2007).
- Martin, C. M. and J. P. Sturmberg (2005) 'General Practice: Chaos, Complexity and Innovation', *Medical Journal of Australia* 183: 106–9.
- Nagarajan, N. and M. Vanheukelen (1997) *Evaluating EU Expenditure Programmes: A Guide*. Brussels: Directorate-general for Budgets of the European Union.

- Patton, M. Q. (1994) 'Developmental Evaluation', *American Journal of Evaluation* 15: 311–19.
- Patton, M. Q. (1997) *Utilization-focused Evaluation* (3rd edn). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2003) 'New Dimensions and Meanings of Useful Evaluation', paper presented at the 2003 conference of the Australasian Evaluation Society, Auckland.
- Pawson, R. (2006) 'Simple Principles for the Evaluation of Complex Programmes', in A. Killoran, C. Swann and M. Kelly (eds) (2006) *Public Health Evidence: Tackling Health Inequalities*, pp. 223–38. Oxford: Oxford University Press.
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: SAGE.
- Perrin, B. (2003) *Implementing the Vision: Addressing the Challenges of Results-Based Management*. Paris: OECD. Available at <http://www.oecd.org/dataoecd/4/10/2497163.pdf>
- Pinnegar, S. (2006) *Are Complex Programs the Best Response to Complex Policy Issues?* City Futures Research Centre Issues paper. Kensington, NSW: University of New South Wales.
- Potter, P. (2004) 'Generalizing from the Unique: Bridging the Gap between Case-Based and Policy-Level Evaluation in the Governance of Neighbourhood Renewal', paper presented at the 2004 Conference of the European Evaluation Society. Available at www.europeanevaluation.org
- Regine, B. and R. Lewin (n.d.) 'Complexity Science in Human Terms: A Relational Model of Business'. Available at <http://www.harvest-associates.com/pubs/human.html>
- Riggan, J. M. (2005) 'Looking Where the Key is: Surfacing Theories of Change in Complex, Collaborative Initiatives', PhD dissertation, University of Pennsylvania. Available at <http://repository.upenn.edu/dissertations/AAI3179796>
- Rogers, P., A. Petrosino, T. Hacsí and T. Huebner (2000) 'Program Theory Evaluation: Practice, Promise and Problems', in P. Rogers, A. Petrosino, T. Hacsí and T. Huebner (eds) *Program Theory Evaluation: Challenges and Opportunities*, pp. 5–13, New Directions in Evaluation series. San Francisco, CA: Jossey-Bass.
- Rogers, P. J. (2000) 'Causal Models in Program Theory Evaluation', in P. J. Rogers, A. J. Petrosino, T. Hacsí and T. A. Huebner (eds) *Program Theory Evaluation: Challenges and Opportunities*, pp. 47–55, New Directions in Evaluation series. San Francisco, CA: Jossey-Bass.
- Rogers, P. J. (2004) 'Logic Models', in S. Mathison (ed.) *Encyclopedia of Evaluation*, pp. 232–4. Newbury Park, CA: SAGE.
- Sanderson, I. (2000) 'Evaluation in Complex Policy Systems', *Evaluation* 6(4): 433–54.
- Schorr, L. (1997) *Common Purpose: Strengthening Families and Neighborhoods to Rebuild America*. New York: Anchor Books Doubleday.
- Sibthorpe, B., N. Glasgow and D. Longstaff (2004) *Complex Adaptive Systems: A Different Way of Thinking about Health Care Systems. A Brief Synopsis of Selected Literature for Initial Work Program – Stream 1*. Canberra: Australian Primary Healthcare Research Institute, Australian National University.
- Smith, M. F. (1989) *Evaluability Assessment: A Practical Approach*. Boston: Kluwer Academic Publishers.
- Sparks, D. (2003) 'NSDC's Theory of Change', Results September 2003. URL: <http://www.nsd.org/library/publications/results/res9-03spar.cfm> (accessed 13 November 2007).
- Stacey, R. (1996) *Complexity and Creativity in Organizations*. San Francisco, CA: Berrett-Koehler.
- Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation* 10(1): 58–76.
- Stufflebeam, D. L. (2004) 'The 21st Century CIPP Model', in M. C. Alkin (ed.) *Evaluation Roots*, pp. 245–66. Newbury Park, CA: SAGE.

Evaluation 14(1)

- Uusikylä, P. and V. Valovirta (2004) 'Three Spheres of Performance Governance: Spanning the Boundaries from Single-Organisation Focus towards a Partnership Network', *Evaluation* 13(4): 399–419.
- Weiss, C. H. (1995) 'Nothing as Practical as a Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families', in J. P. Connell, A. C. Kubisch, L. B. Schorr and C. H. Weiss (eds) *New Approaches to Evaluating Community Initiatives*, pp. 65–9. Washington, DC: Aspen Institute.
- Weiss, C. (1998) *Evaluation: Methods for Studying Programs and Policies*. Englewood Cliffs, NJ: Prentice Hall.
- Wholey, Joseph S. (2003) 'Using Policy Evaluation to Support Decision Making and Improve Program Performance', *Japanese Journal of Evaluation Studies* 3(2): 6–21.
- Winston, J. (1999) 'Understanding Performance Measurement: A Response to Perrin', *American Journal of Evaluation* 20(1): 95–100.
- W. K. Kellogg Foundation (2004) *Logic Model Development Guide*. Battle Creek, MI: W. K. Kellogg Foundation. Available at <http://www.wkfk.org/Pubs/Tools/Evaluation/Pub3669.pdf>
- World Bank (2004) 'Pages from World Bank History: The Fight Against Riverblindness'. Available at <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/EXTARCHIVES/0,,contentMDK:20098846~pagePK:36726~piPK:36092~theSitePK:29506,00.html>

PATRICIA J. ROGERS is Associate Professor of Public Sector Evaluation at the Royal Melbourne Institute of Technology, with a particular interest in evidence-based policy and practice in complex human service fields. Please address correspondence to: Centre for Applied Social Research, RMIT University, 124 Latrobe Street, Melbourne, VIC 3000, Australia.
[email: Patricia.Rogers@rmit.edu.au]