

Criteria for assessing the evaluability of a Theory of Change

Rick Davies, Thursday, 05 April 2012. Posted on "[Rick on the Road](#)"

Our team has recently begun work on an evaluability assessment of an agency's work in a particular policy area, covering many programs in many countries. Part of our brief is to examine the evaluability of the programs' Theory of Change (ToC).

In order to do this we clearly need to identify some criteria for assessing the evaluability of ToC. I initially identified five which I thought might be appropriate, and then put these out to the members of the [MandE NEWS email list](#) for comment. Many comments were quickly forthcoming. In all, a total of 20 people responded in the space of two days (Thanks to Bali, Dwiagus, Denis, Bob, Helene, Mustapha, Justine, Claude, Alex, Alatunji, Isabel, Sven, Irene, Francis, Erik, Dinesh, Rebecca, John, Rajan and Nick).

Caveats and clarifications

What I have presented below is my current perspective on the issue of evaluability criteria, as informed by these responses. It is not intended to be an objective and representative description of the responses ([Look here for a copy of all the comments received](#))

The word evaluable needs some clarification. In the literature on evaluability assessments it has two meanings. The main one is that it is *possible* to evaluate something. For example, if the theory is clear and the data is available. The second meaning is more *practically oriented*. The theory may be clear and the data available, but the theory may be so implausible that it is simply not worth expending resources on its evaluation. Or there may be a perfectly good ToC, but if no one owns it apart from a consultant who visited the project six months ago, so it might be questionable whether expensive resources should be invested in its evaluation.

We also need to distinguish between an evaluable ToC and a "good" ToC. A ToC may be evaluable because the theory is clear and plausible, and relevant data is available. But as the program is implemented, or following its evaluation, it might be discovered that the ToC was wrong, that people or institutions don't work the way the theory expected them to do so. It was a "bad" ToC. Alternately it is also possible that a ToC may turn out to be good, but the poor way it was initially expressed made it un-evaluable, until remedial changes were made.

This brings us to a third clarification. My minimalist definition of a ToC is quite simple: "*the description of a sequence of events that is expected to lead to a particular desired outcome*" Such a description could be in text, tables, diagrams or a combination of these. Falling within the scope of this definition we could of course find ToC that are evaluable and those that are not so evaluable.

A possible list of criteria for assessing the evaluability of a Theory of Change (Version 2)

- Understandable
 - Do the individual readers of the ToC find it easy to understand? Is the text understandable? If used, is the diagram clear?
 - Do different people interpret the ToC in the same way?

- Do different documents give consistent representations of the same ToC?
- Verifiable
 - Are the events described in a way that could be verified? This is the same territory as that of Objectively Verifiable Indicators (OVIs) and Means of Verification (MoVs) found in LogFrames
- Testable
 - Are there identifiable causal links between the events? Often there are not
 - Are the linked events parts of an identifiable causal pathway?
- Explained
 - Are there explanations of how the connections are expected to work? Connections are common, explanations of the causal process involved are much less so.
 - Have the underlying assumptions been made explicit? (also duplicated below)
- Complete
 - Does what might be a long chain of events make a connection between the intervening agent with the intended beneficiaries (/target of their actions)? In a recent ToC that I have seen the ToC is quite detailed at the beneficiary end, but surprisingly vague and unspecific towards the agent's end, even though that is where accountability might be more immediately expected.
- Inclusive (a better a term is needed here)
 - Does the ToC encompass the diversity of contexts it is meant to cover? In ToC covering whole portfolios of projects there could be a substantial diversity of contexts and interventions. Does the ToC provide room for these with sacrificing too much in terms of verifiability and testability" See [Modular Theories of Change: A means of coping with diversity and change?](#) for some views how to respond to this challenge.
- Justifiable(new)
 - Is there evidence supporting the sequence of events in the ToC? Either from past studies, previous projects, and/or from a situation analysis/baseline study or the like which is part of the design/inception stage of the current project
- Plausible (new)
 - Where there is no prior evidence is the sequence of events plausible, given what is known about the intervention and the context?
 - Have the underlying assumptions been made explicit?
 - Have contextual factors been recognised as important mediating variables?
- Owned
 - Can those responsible for contents of the ToC be identified?
 - How widely owned is the ToC?
 - Do their views have any consequences?
- Embedded
 - Are the contents of the ToC are also referred to in other documents that will help ensure that it is operationalized?

Weighting

It was sensibly suggested that some criteria were more important than others. One argued that if you can establish that the causal links in a ToC are evidence based then ‘ownership will and shall follow’.

In individual evaluability assessments a simple sense of their relative priority may be sufficient. When comparisons need to be made of the evaluability of multiple programs, it may be necessary to think about weighted scoring mechanisms/checklists.

Purpose

It was suggested that the criteria used would depend on the purpose for which the ToC was created. An understanding of the Purpose could therefore inform the weighting given to the different criteria.

Prior to consulting the email list members I had drafted a list of three possible purposes that could generate different kinds of evaluation questions, which an evaluability assessment would need to consider. They were:

- If the purpose of the ToC was to **set direction**
 - Then we need to ask were programs designed accordingly?
- If the purpose of the ToC was to **make a prediction**
 - Then we need to ask if the programs subsequently turn out this way
- If the purpose of the ToC was to **provide a summation**
 - Then we need to ask if this is an accurate picture of what actually happened?

One criticism of the inclusion of prediction was that most ToC are nothing like scientific models and because of this they are typically insufficient in their contents to generate any attributable predictions. This may be true in the sense that scientific predictions aim to be generalisable, albeit subject to specific conditions e.g. that gravity behaves the same way in different parts of the universe. But most program ToC have much more location-specific predictions in mind, e.g. about the effects of a particular intervention in a particular place. There are interesting exceptions however, such as a ToC about a whole portfolio of programs, or a ToC about a whole policy area that might be operationalised through investment portfolios managed in a range of countries. There the criticism of incapacity may be more relevant.

The same critic proposed an alternate purpose to prediction, one where simplicity might be more of a virtue than a liability. ToC may aim to communicate or generate **insight**, by focusing on the core of an idea that is driving or inspiring a program. If so, then evaluation question could focus on how the ToC has changed the users’ *understanding* of the issues involved. This question about effects could be extended to include the effects of participation in the process whereby the ToC was developed.

Issues arising about criteria

The following issues were raised.

- **Process and Product:** The list above is largely about the ToC product, not the process whereby it was created. Some argued there needed to be a participatory process of development to ensure the ToC was “*aligned with the needs of beneficiaries and the national objectives*”. However, others argued that that “*ToC are not “development projects” that must be aligned with the Paris*

Declaration, but rather tools that must be rigorous, applied without 'complaisance' ". The hoped for reality might lie in between, ToC typically are associated with specifically project interventions and the extent of their ownership is relevant to answering the practical aspects of evaluability. On the other hand, the rigour of their use as tools will affect their usefulness and whether they can be evaluated. The product-oriented criteria given above do include two criteria that may reflect the effects of a good development process. i.e. ownership and embeddedness.

- **Ownership:** It was argued that ownership was not a criterion of good ToC, often the consensus in science has been proved wrong. But in the above list the criterion of ownership is relevant to whether the ToC is worth evaluating, it is not a criterion of value of the belief or understanding represented by the ToC. It could be argued that widely owned views of how a project is working are eminently worth evaluating, because of the risk that they are wrong.
- This approach might lead to the view that on the other hand ToC with few owners should not be evaluated. This view was in effect questioned by an example cited of an evaluator coming up with their alternative ToC, which was based on prior evaluations studies and research, in contrast to the politically motivated views of the official in charge of a program. This brings us back to the criteria listed above, and the idea of weighting them according to context (ownership versus justifiability).
- **Relevance:** This proposed criterion begs the question of relevant to whom? Ownership of the ToC (voluntary or mandated) would seem to signify a degree of relevance.
- **Falsifiability:** It was argued that this is the pre-eminent criteria of a good scientific theory, and one which needed more attention by development agencies when thinking about the ToC behind their interventions. The criteria in the list above address this to some extent by inquiring about the existence of clear causal links, along with good explanations for how they are expected to work. Perhaps "good" needs to be replaced by falsifiable, though I worry about setting the bar too high when most ToC I see barely manage to crawl. Many decent ToC do include multiple causal links. The more there are, the more vulnerable they are to disproof, because only one link needs to fail for ToC not to work. This could be seen as a crude measure of falsifiability.
- **Flexibility:** Although it was suggested that ToC be flexible and adaptable this view is contentious, in that it seems to contradict the need for specificity (by being verifiable, testable, and explained) and thus its falsifiability. However, there is no in principle reason why a ToC can't be changed. If it is, it becomes a different ToC, subject to a separate evaluation. It is not the same one as before. The only point to note here is that the findings of the adapted version would not validate the content of the earlier version.
- Lack of adaptability may also be a problem. It was suggested evaluators should ask 'When has the ToC been reviewed and how has it been adapted in the light of implementation experience, M&E data, dialogue and consultation with stakeholders?' If the answer is not for a long time, then there may be doubts about its current relevance, which could be reflected in limited ownership.
- **Clarity of logic as well as evidence:** One commentator suggested that it might be made clear whether a given cause is both "necessary and sufficient", presumably as distinct from alternative

combinations of these terms. Necessity *and* sufficiency is a demanding criterion, and arguable whether which many programs would satisfy, or perhaps even should satisfy.

- **Simplicity:** This suggested requirement (captured by Occam's razor) is not as simple a requirement as it might sound. It will always be in tension with its opposite (captured by Ashby's Law of Requisite Variety), which is that a theory must also have sufficient internal complexity in order to describe the complexity of the events it is seeking to describe. Along the same lines some commentators asked whether there was enough detail provided, the lack of which can affect verifiability and testability. Simplicity may win out as the more important criteria where a ToC is primarily intended as a communication tool.
- **Justifiability** was highlighted as important. **Plausibility** was questioned "*What that does really mean? If based on common sense then it is incompatible with being evidence based! If humanity had to rely on common sense, the earth would still be flat!!*" Plausibility is clearly not a good evaluation finding. But it is a useful finding for an evaluability assessment. If a ToC is not plausible then it makes no sense to go any further with the design of an evaluation. Justifiability is evidence of a good ToC, and is a judgement that might follow an evaluation. However, it might also be obvious before an evaluation, through an evaluability assessment, and lead to a decision that a further evaluation would not be useful.

Informed sources mentioned by contributors

Connell, J.P. & Kubisch, A.C. (1998) *Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects and problems*, in: K. Fulbright-Anderson, A.C. Kubisch & J.P. Connell (Eds) *New Approaches to Evaluating Community Initiatives. Volume 2: Theory, measurement and analysis* (Queenstown, The Aspen Institute). [courtesy of [John Mayne](#)]

Connell and Kubisch suggest a number of attributes of a good theory of change.

- It should be **plausible**. Does common sense or prior evidence suggest that the activities, if implemented, will lead to desired results?
- It should be **agreed**. Is there reasonable agreement with the theory of change as postulated?
- It should be **embedded**. Is the theory of change embedded in a broader social and economic context, where other factors and risks likely to influence the desired results are identified?
- It should be **testable**. Is the theory of change specific enough to measure its assumptions in credible and useful ways?

Kusters, C., van Vugt, S., Wigboldus, S., Williams, B., Woodhill, J. (2011) [Making Evaluations Matter: A Practical Guide for Evaluators](#). Centre for Development Innovation, Wageningen University & Research centre, Wageningen. [Courtesy of Awolaye Olatunji]

Under the heading "Articulate the Theory of Change" they ask:

- "What is the envisaged pathway of change for the development intervention?"
- How do stakeholders think change will happen?
- What critical stakeholders think change will happen?

- What critical assumptions have been made for assumptions have been made for this pathway of change?