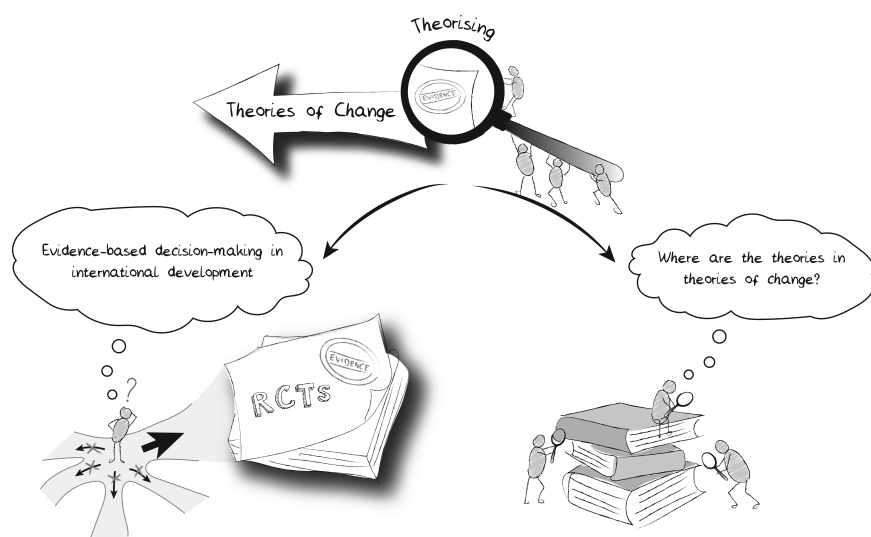


10

THEORISING THEORIES OF CHANGE IN INTERNATIONAL DEVELOPMENT

What counts as evidence?

Katarzyna Cieslik and Cees Leeuwis



Introduction

Development projects are based on explicit or implicit assumptions of how and why a certain intervention will work. Over the past three decades, these assumptions have captured the attention of academics and practitioners alike, sparking interest in the so-called *theories of change* (Brown, 2020; Prinsen and Nijhof, 2015). In the most common sense, theories of change (ToCs) are explanations of how groups of stakeholders expect to reach a commonly understood long-term goal (Brest, 2010; Stein and Valters, 2012; Weiss, 1995). Depending on

their purpose (designing, monitoring and evaluation, or scaling interventions), theories of change might be developed and deployed at the level of a singular intervention, a complex program, or an entire organisation (Mason and Barnes, 2007). Currently, many development donors demand that projects and initiatives are based on an explicit and credible theory of change, turning it into a widely used planning and assessment tool in international development.

Interestingly, to date, the relationship between theories of change (as tools in development practice) and social scientific theories *about* change has not been investigated in academic literature. In most cases, theories of change rely on the common-sense logic of the implementers and on experimental evaluations of similar interventions (Brest, 2010; Astbury and Leeuw, 2010), referred to as 'evidence.' As evidence-based decision-making has become the cornerstone of sustainable development, so have theories of change; both promising to replace ideologically driven policy with rational planning (Donovan, 2018).

In this paper, we explore the links between theories of change and scientific knowledge, theories, and methods. In the context of international development, a scientific theory would describe the 'how' and 'why' of achieving positive social change. It should also fulfil two conditions: adherence to a scientific method, and a certain level of generalisability. At the same time, in lay terms, a 'theory' is understood as a hypothesis, or a 'hunch,' that has little to do with science. By scrutinising how these 'hunches' and 'evidence' both guide ToC development, we critically assess the use of knowledge and science in the design, monitoring and assessment of development interventions. Our contribution is as follows.

First, we establish conceptual clarity over the existing conceptualisations of ToCs, positioning them against similar approaches: Logical framework approaches (LFAs) and program logic models (PLMs). Broadly categorised as development management 'tools,' all of these approaches were driven by the quest for increased efficacy and accountability within the development sector. We argue that while LFAs and PLMs belong to the realm of development management, strategic planning, and evaluation, ToCs originated from the discipline of critical development studies.

Second, we show that the use of 'evidence' in ToCs development rarely goes beyond looking at existing assessments of similar interventions. These are generated through randomised controlled trials (RCTs) and systematic research syntheses. We argue that experimental evaluations ('evidence'), while useful to policy makers, do not, in fact, advance theory formation, as they do not consider the auxiliary assumptions and their results are not generalisable. As such, they can guide policy makers to evaluate and assess social interventions, but they do not substantially further our understanding of the outside world, nor do they capitalise on validated scientific theories about how change happened in the past.

Third, we introduce alternative sources of 'evidence' that should be considered complementary to positivist approaches in pursuit of improved decision-making. Instead of discarding 'hunches' and 'common-sense logic' as unscientific, we propose broadening the spectrum of perspectives engaged in ToC development

and use. Drawing on the tradition of participatory development, we suggest that inclusive project governance can lead to more robust ToCs based on contextual knowledge. While including local-level stakeholders' perspectives results in more locally relevant ToCs, including social science theories *about* change (the 'how' and 'why') allows insight into potential auxiliary assumptions.

With this, we argue that experimental assessments are not the only way in which science may contribute to 'theories of change' in international development. We urge policy makers to look beyond what is considered 'hard evidence,' carving space for more qualitative, inclusive, and deliberative approaches that highlight the 'why' and 'how' of the change process (Murdach, 2010). In so doing, we undermine the positivist notion of 'evidence-based' policy, arguing instead for broader 'evidence-informed' decision-making. With these contributions, we complement the existing literature on the role of scientific knowledge in international development with a nuanced analysis of the interplay of different functions of theory building and their evolving legitimacy within the legacy of development studies.

Evidence-based decision-making in international development

In the name of accountability

The 2019 Nobel Prize in Economic Sciences, awarded to Abhijit Banerjee, Esther and Michael Kremer 'for their experimental approach to alleviating global poverty,' put 'evidence-based' policy in the spotlight. Their Abdul Latif Jameel Poverty Action Lab (J-PAL) has been conducting rigorous assessments of social interventions through randomised experiments in developing countries for almost 20 years, and building a 'theory of change' is the first step in their approach. To J-PAL, theory of change provides a structured way of 'thinking about impact,' integrating program design, implementation, monitoring and evaluation, and communication. The concerted movement to enhance development effectiveness through better planning and increased accountability builds on the assumption that, through the application of experimental evaluations, the development industry can rid itself of ineffective interventions and scale-up those that do work. Apart from J-PAL, a plethora of similar organisations started conducting or commissioning RCTs and systematic syntheses of them, e.g. the International Initiative for Impact Evaluation (3ie) and Innovations for Poverty Action, aiming to quantitatively prove what works, for whom, why, and at what cost in low-and middle-income countries (Pahlman, 2014).

The growing popularity of evidence-provisioning organisations like J-PAL runs parallel to increasing transparency and accountability pressures within the international development sector, which Ramalingam et al. (2014, p. 3) call an 'accountability revolution.' Similarly, the transformations within the aid industry—including the economisation of development and the retreat from macro-planning, created the opportunity for the 'rise of the randomistas,' or proponents

of applying experimental evaluations to social program in development contexts (Donovan, 2018, p. 27; Leigh, 2018). Delivering development results has become a prominent topic within the development industry. Referred to as ‘the golden age of evidence-based policy’ (APPAM, 2015) or ‘the quiet movement to make government fail less often’ (Leonhard, 2014), the new millennium was marked by an interest in better planning, strategic management, and continuous, rigorous assessment of projects, interventions, and policies (de Souza Leão and Eyal, 2019).

Importantly, as noted by Baguios (2019), the increased use of planning and of assessment tools such as theories of change was not meant to increase accountability towards the aid receivers, but to provide clarity and reassurance to the donors. Brest (2010, p. 47) explains that ‘[...] a funder has a legitimate interest in knowing whether an organization is on the path to success and, at some point, whether it is actually achieving impact.’ Accordingly, experimental assessments gain popularity as authoritative means of achieving certainty over ‘what works.’ For RCTs to take place, however, the relationship between the intervention and the desired outcome needs to be explicated. This is where logframes, logic models, and theories of change come in. In the next two sections, we first look how logframes evolved into theories of change, and then scrutinise the ‘evidence’ that they are meant to build on.

Logframes, logic models, and theories of change

Chronologically, logframes and logic models much preceded theories of change. Broadly speaking, logframes, and the management approach that followed (the logical framework approach, or LFA) are a program design methodology that delineates the core elements of an intervention (see Figure 10.1):



FIGURE 10.1. Logical framework: core components, from ‘inputs’ to ‘impacts’

The core elements are inputs (the resources, contributions, and investments that go into a program), outputs (the visible and tangible consequences of program project input), outcomes (the short-term effects of the program/project), and impacts (long-term, generalised goals) (Gasper, 2000). When portrayed in a 4x4 matrix, logframe becomes a ‘logic model’ or program logic model (PLM). Logic models are hypothesised chains of causes and effects, leading to a generalised long-term goal, often taking the form of ‘if-then’ relationships between the core elements. In this sense, LFA and PLM both include implicit or explicit assumptions regarding the contextual conditions in which interventions take place (see Figure 10.2). In addition, both tools enable monitoring and evaluation process by specifying quantifiable indicators for each of the implementation stages (Gasper, 2000).

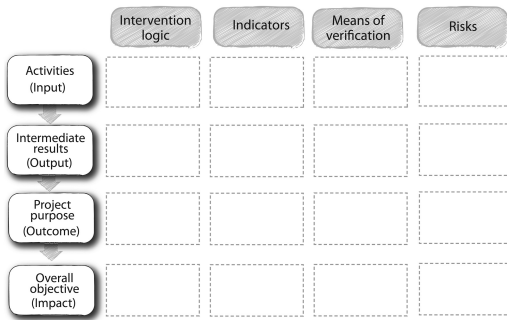


FIGURE 10.2 Logic model known as a 4x4 matrix

As an entry point to structured evaluations, LFA/PLM are often seen as a way to plausibly demonstrate impact, and hence of increasing transparency and accountability of the development industry. At the same time, they impose a pre-defined project logic on stakeholders with potentially different—or even conflicting—worldviews. A number of development organisations refer to their logframes as ‘roadmaps’ or ‘blueprints,’ discounting any alternative trajectories that their interventions might have (Dale, 2003). As narrative framing tools, LFA/PLM are also target-driven and invariably positive (Büscher, 2014) (Table 10.1).

From the point of view of social science theory, LFA/PLM reflects a strong belief in rational planning and control as a mechanism for orchestrating positive change, as it assumes that change and transformation can be engineered, predicted, and designed by means of systematic thinking about hierarchies of goals and means (inputs, outputs, outcomes, impacts). However, numerous studies in development sociology suggest that development and change arise from a capricious process of social struggle over resources, meanings, goals, and identities with inherently uncertain outcomes (Leeuwis, 2000). Similarly, historians studying transformation over longer periods have concluded that meaningful change emerges from competition between those supporting and those challenging the status quo, with success depending on the quality of coalition formation and adaptive learning in the context of ever changing circumstances (Klerkx et al., 2010). Thus, we see that logic models may be informed by modes of thinking at a more abstract level (e.g. on whether change emerges from ‘planning,’ ‘learning,’ or ‘social struggle’; see Leeuwis, 2000) allowing us to question whether a ‘rational planning and control’ paradigm is an adequate reflection of how change occurs.

Specifically, from within the discipline of development studies came a strong critical narrative, juxtaposing LFA/PLM and bottom-up inclusive approaches like Participatory Rural Appraisal (PRA) and Participatory Learning and Action (PLA) (Aune, 2000; Kumar and Corbridge, 2002). As a response, theories of change (ToC) were meant to bridge structured planning and local participation

TABLE 10.1 The three uses of project planning tools: overview of main criticisms

Three main uses of LFA/PLM	Criticisms
Designing interventions	• Fail to consider historical and social science theories about how change happened in the past. • Based on ‘expert hunches’ and, at best, some of the existing ‘evidence’ (RCTs). • Promote a singular project trajectory, disregard alternative pathways. • Hamper innovation and creativity at the local level.
Monitoring and evaluation	• Create implementation bottle necks, slow down project maturation. • Discount externalities (‘unintended effects’).
Scaling up and out	• One-fits-all solution (‘blueprint’ approach). • Disempower local stakeholders (a ‘straightjacket’).

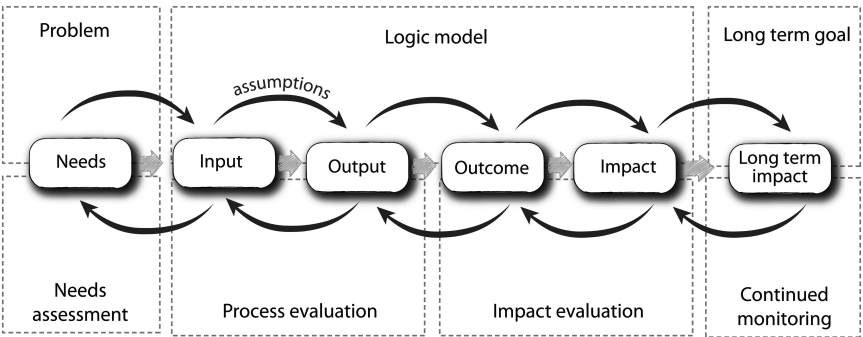


FIGURE 10.3 Basic setup of a theory of change

(Weiss, 1995). Though the principal components of ToCs resemble LFA/PLM, their originality lies in recognising the importance of tacit, or only partly articulated assumptions, of how and why an intervention is supposed to work (Astbury and Leeuw, 2010; Weiss, 1995). As Figure 10.3 illustrates, ToCs start with problem framing (‘needs assessment’) which influences the course(s) of action taken (‘assumptions’ towards input-output-outcome-impact relationships). While within a limited program-timeframe of several years a ToC would usually end with an ‘impact evaluation,’ further ‘assumptions’ need to be elaborated to explain how and why the project will achieve lasting change (‘long-term goal’).

By involving the different level stakeholders, ToCs encourage local organisations to take ownership of, and the responsibility for, the course of the interventions (Sullivan and Stewart, 2006). This requires sensitivity to perennial power imbalances within project structures, which in reality is rarely achieved

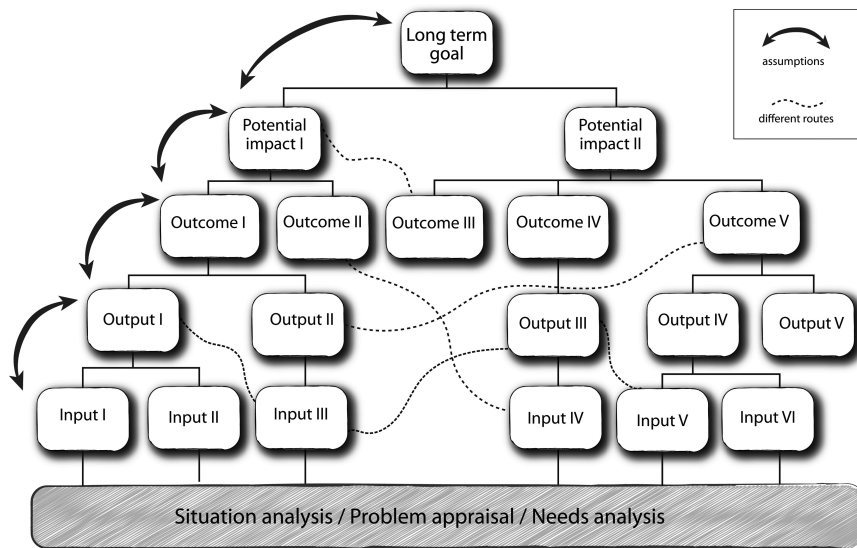


FIGURE 10.4 Theory of change template graph

(Ferguson et al., 2010). As Figure 10.4 illustrates, ToCs acknowledge that relationships between inputs and outputs are not straightforward and may take different routes (see dotted lines on the graph).

The brief reconstruction of the evolution of logical frameworks, logic models, and theories of change points to the constant tension between stability and flexibility in project management (Hersoug, 1996). While development studies scholars argue that standardisation of the project approach is detrimental to local agency, constrains innovation, and is prone to ignore contextual dynamics of power and competition, development managers point out that the lack of structure confuses evaluation and hampers scale-up efforts (Crawford and Bryce, 2003; Curtis and Poon, 2009).

While adopting any of the LFA, PLM, and ToC tools may help development practitioners attenuate some of the projects' operational uncertainties, it does nothing for contextual uncertainty (Hersoug, 1996). When trying to decide what kind of intervention is most likely to lead to the desired long-term goal and in mapping out the project trajectory, founders, managers, and implementers turn to research to provide them with 'evidence' of 'what works.' In order to avoid bias and formulating 'unfounded' assumptions, ToCs are meant to be 'evidence-based.' In the sections that follow, we take a critical look at what kind of scientific outputs are used to construct theories of change for development interventions.

TABLE 10.2 The three uses of theories of change: comparing disciplinary perspectives

<i>Three main uses of LFA, PLM, ToC, and PoC</i>	<i>Framing by development management</i>	<i>Framing by critical development studies</i>
Designing interventions	Action plans, roadmaps, blueprints.	Participatory needs assessment, co-creation, co-design, negotiating differences.
Monitoring and evaluation	Results-based management, performance review, cost-benefit analysis.	Reflexive critique and feedback loops, participatory evaluation, flexibility, adaptation to take in unforeseen surprises.
Scaling up and out	Impact acceleration, value maximisation.	Learning, context-specificity, communities of practice.

What counts as evidence? Hierarchies of evidence and the 'gold standard' of RCTs

While 'evidence-based policy' has become the dominant paradigm in policy circles, it is often performed with perfunctory attention. In a comprehensive report commissioned by the King's Fund, UK, Coote et al. (2004, p. xi) find that 'major social programs,' were not, in fact based on rigorous evidence: 'Interviews with those in central government make it clear that they (social programs) have been designed, by and large, on the basis of informed guesswork and expert hunches, enriched by some evidence and driven by political and other imperatives'; see also Mason and Barnes (2007).

The issue of what constitutes 'evidence' for research-informed policy triggered considerable academic debates (Oakley, 2000). When applied to theories of change, however, the answer appears much more straightforward: 'evidence' refers to evaluations of similar interventions and proofs that a certain logic 'works' (Pahlman, 2014). While, taken in its entirety, evaluation can generate many kinds of knowledge, not all knowledge is routinely defined as evidence: there is a clear tendency to put systematic reviews of RCT studies at the top of the hierarchy, followed by single RCT studies and quantitative survey results. Qualitative research (e.g. ethnographic studies, expert interviews) may also be used (Coote et al., 2004), though it would count as 'soft evidence' (Murdach, 2010; Oakley, 2000). Evidence that comes low in such hierarchies is likely to be ignored, including local knowledges and lived experience documented through ethnography and anthropology.

RCTs are a type of impact evaluation that randomises access to a particular social intervention to produce an unbiased and internally valid impact estimate. There are a number of reasons why RCTs as scientific outputs rank highly in policy makers' hierarchies of evidence. First, RCTs demonstrate a clear causal relationship between an intervention and its outcomes. This is

because randomisation eliminates the bias inherent in comparative studies: as all potentially impactful variables are equally distributed in treatment and control groups, any possible difference in outcome can only be attributed to the intervention. Second, RCTs allow for the quantification of uncertainty. If the sample that undergoes randomisation is large enough, RCTs provide a quantifiable degree of certainty about the accuracy of the captured effect. Third, by comparing more than one treatment, RCTs allow researchers to determine which components of a program (or, in ToC language, which ‘inputs’) are necessary for it to be effective in a *ceteris paribus* context. Fourth, they allow for a thorough cost-benefit analysis, by setting the price (cost of activities, or inputs) of achieving incremental change in a chosen indicator (quantifiable outcome and/or impact).

These four features make RCTs powerful tools for evaluation. At the same time, RCTs do exactly what they are designed to do: they assess single projects, in particular contexts. In other words: though a highly effective evaluation methodology, RCTs do not show ‘what works,’ but ‘what was observed to work under specific circumstances,’ without saying anything about these circumstances. Against this background, and considering that RCTs and RCT-syntheses are almost exclusively used as base for ‘evidence-based policy,’ it is crucial to determine what contribution RCTs actually make to theory-building.

Where are the theories in theories of change?

RCTs as means to test theories

A scientific theory is an explanation of a chosen aspect of the world that can be repeatedly tested using accepted protocols of observation, measurement, and evaluation of results. Just as laboratory experiments are set up to test scientific theories, RCTs are meant to test program theories: to establish whether a certain intervention brings about the desired effect. In laboratory conditions, each testing of a theory entails ‘auxiliary assumptions,’ which are all the propositions that are assumed to be accurate or in place in order for the test to work, such as a particular temperature or humidity as measured by certain devices in a biophysical experiment. Field experiments, on the other hand, do not, and cannot, consider auxiliary assumptions, as it is impossible to enlist all the specifications of social reality (Cartwright, 2007). To the contrary, as rigorous evaluation tools, RCTs are meant to detect the effect of an intervention *regardless* of the external conditions, such as the unique socio-cultural characteristics of a given region. Thanks to randomisation, these will affect the control and treatment groups in exactly the same way, and hence can be ignored for the purpose of assessment. They do matter, though, for external validity. This, in fact, is the function of the counterfactual: as long as the control and treatment groups are exposed to the same, undetermined set of influences, these influences do not matter, as the experiment will only consider the difference in outcome. A positive result of a

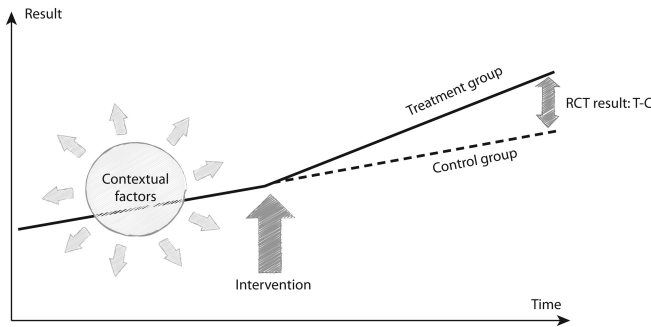


FIGURE 10.5 A schematic representation of an RCT set-up and outcome

rigorously executed RCT will always detect the appropriate causal conclusion—and only that (Figure 10.5).

While strikingly effective for evaluations, this feature of RCTs translates into a serious flaw when assessing their theory-building potential. RCT results are limited to the exact temporal and special context where they were conducted, and can never be generalised. Cartwright (2007, p. 11) writes that ‘the benefit that the conclusions follow deductively in the ideal case comes with a great cost: narrowness of scope. This is an instance of the familiar trade-off between internal and external validity. RCTs have high internal validity but the formal methodology puts severe constraints on the assumptions a target population must meet to justify exporting a conclusion from the test population to the target.’ When applied to social science research, the absence of auxiliary assumptions in RCTs renders generalisation of results virtually impossible. Accordingly, using experimental assessment of one intervention as ‘evidence’ to support another is methodologically unsound. In addition, while RCTs detect causality, they remain a ‘black box’ in terms of the ways by which the intervention exerted its impact.

Against this background, the relationship between theories of change and randomised controlled trials is problematic at best. First, program theories are tested through RCTs: in the case of a positive assessment, a ToC is considered to ‘have worked.’ Second, in accordance with the evidence-based policy paradigm, these results are used to guide the development of ToCs for future interventions, in new places, and by different actors. Both inferences are questionable to some degree. While the first may be appropriate in the sense that RCTs can provide robust and rigorous assessment that an intervention has worked, it still remains possible that the intervention has worked in a different way than its theory of change suggests. For example, a theory of change might state that community meetings to promote technology create awareness of the positive consequences of technology use (in line with scientific theories about diffusion of innovations), while in practice they may

have served to resolve conflict or promote collective action in support of technology uptake (in line with systemic theories about innovation). Thus, a positive assessment of the intervention provides no definitive evidence that the explicated theory of change was in fact correct. As explained above, the second inference—that the intervention is likely to lead to similar effects elsewhere—is even more questionable, as RCT results cannot be generalised.

Importantly, as discussed above, theories of change are supposed to ‘spell out’ project propositions and make implicit assumptions explicit. At the same time, without proper tools and procedures of capturing the ‘auxiliary assumptions,’ these can only comprise what Coote et al. (2004, p. xi). call ‘informed guesswork and expert hunches.’ In the next section, we propose what these tools and procedures might be, and develop a more comprehensive evidence–policy model.

Theorising theories of change: broadening out and opening up

In the practitioner and grey literature cited in this paper, evidence-based theories of change are presented as the ‘gold standard’ and other approaches are often dismissed. At the same time, experience-based insights of the local staff are also not considered: ‘a common element across professions is the extent to which the legitimacy of professional decision-making is no longer based on what might be accounted as professional wisdom, often founded in tradition. Instead, it is thought to reside in the weight of evidence, produced by other members of the community or by the researcher community, independently sifted through external review’ (Clegg, 2007, p. 417). This singular logic mindset and expert supremacy, however, is precisely the problem that the theories of change were meant to address by engaging with multiple-level stakeholders (Sullivan and Stewart, 2006). Once an intervention is considered ‘successful’—often with the help of an RCT—its theory of change is taken out of its original context and implemented in another through scaling up; losing both its place specificity and its participatory credentials. Against this background, we propose that the original, stakeholder-driven theory of change model should not only be reinstated, but also further opened up and broadened out to include both more local stakeholders, and a wider range of interdisciplinary scholarship.

In defence of common sense

The ‘opening up’ entails inviting different level stakeholders to participate in the theory of change process, besides policy makers and practitioners. In a critical piece on ToCs uses and misuses, Mason and Barnes (2007, p. 162) ask rhetorically, ‘should we assume that a ToC generated by social scientists on the basis of a review of research evidence is necessarily better than that produced by staff and community members directly involved in the program itself?’ further observing that ‘if services (within international development organizations) are

commissioned solely on the basis of research evidence, there may be little space for innovation.’ Local communities and frontline staff may be better positioned to judge or make assumptions about where and how a ToC needs to be tweaked in order to achieve a better fit with the local context, even though it is based on ‘common-sense logic’ rather than research evidence. This applies to ToC development, but also to deployment (on-the-go adaptation of project activities and thresholds) and to evaluation (designing locally relevant indicators of project ‘success’ which may be different from the onset assumptions). While concerned communities are at times participating in the ToC development workshops, at later stages their engagement is usually limited to ‘consulting’ and ‘informing.’ This is at odds with the central idea of participatory development practice, where a broad, public, deliberative conversation is essential for reaching a shared understanding of the problems and monitoring the process of change over time. Central to the process is creating a shared space through transparency and openness, and actively resisting the systemic power pressures. Apart from content-enrichment and enhanced insight into locally relevant conditions, including more local voices would contribute to what Baguios (2019) calls the ‘decolonization of the project management in the aid sector.’ At the same time, romanticising and glorifying ‘local voices’ bears the risk of raising unrealistic expectations as well as losing sight of the ‘big picture’: the organisational mission.

Carving space for social science theories about change

Broadening the range of social science approaches to ToC development requires undermining the hierarchy of ‘research evidence.’ As indicated earlier, evidence from RCTs falls short on the question of ‘how and why’ change happens. There is a plethora of historical, sociological, and anthropological meta-level theories that shed light on processes that matter in change trajectories (e.g. complexity theories of societal change, progressive change theory, or innovation theory). Although such theories tend to be validated through methodological approaches other than RCTs (e.g. systematic process tracing, historiography, discourse analysis, participant observation, etc.) they may be usefully introduced in discussions with stakeholders for purposes of ToC construction, adaptation, and testing. For example, if social science theories suggest that meaningful transformation depends on the quality of coalition formation and adaptive learning, then such a theory could enrich the stakeholder consultations by questioning whether suggested ‘inputs, outputs, and outcomes’ in the theory of change (and indicators used to assess these) indeed refer to such parameters at all.

For this reason, we argue that applying a range of social scientific approaches to ToC development, implementation, and evaluation ultimately leads to better projects based on well-grounded theories and assumptions about how change happens. Different theoretical traditions have complementary insights and functions, allowing practitioners and scholars to challenge the conventions and innovate. Combining experimental assessment with a comprehensive and inclusive

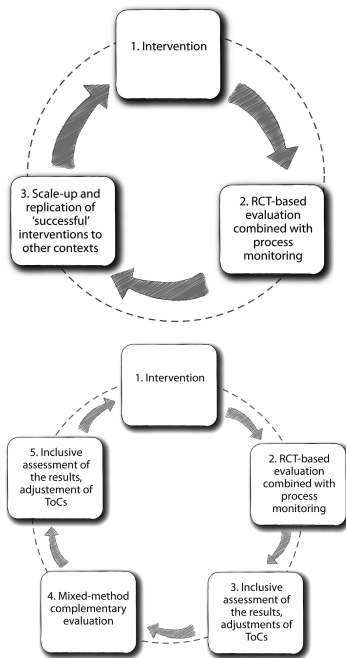


FIGURE 10.6 Currently dominant (a) and proposed (b) use of methods in ToC evaluation cycles

local participation leads to both more locally relevant projects, and more meaningful evaluations. Conceived as such, participatory assessment of the RCT results could help the researchers uncover the contextual factors—or the auxiliary assumptions—that could have contributed to the intervention’s success/failure results, as well as shed light on the change process (the ‘why’ and ‘how’, see Figures 10.6a and 10.6b).

Conclusions

In the recently published *Navigation by Judgment Why and When Top Down Management of Foreign Aid Doesn’t Work*, Honig (2019) provides a strong critique of development management tools and donor control. Drawing on a database of over 14,000 development projects, he argues that a narrow focus on reaching pre-set targets limits frontline workers from solving problems on the ground, seriously undermining the projects’ performance. Conscious of these critiques, in this paper, we traced the evolution of logic models, logframes and theories of change, highlighting the trade-offs between transparency and accountability on one hand, and flexibility and adaptation on the other. We showed how the

growing popularity of theories of change is linked to different concerns regarding accountability within the development sector.

While ToC was initially developed as a method to induce greater accountability to local-level stakeholders with appropriate knowledge and understanding of relevant conditions, we now witness that they are increasingly used as a point of reference in the development and identification of 'evidence-based' policies and interventions that can be generalised across contexts. We critically discussed the concept of evidence-based policy making, arguing that focusing solely on experimental methods (RCTs) as evidence sources is not just epistemologically limiting, but also methodologically unsound. The relationship between policy and evidence is far more complex than a linear evidence-to-policy chain suggests. Central to making evidence more meaningful to the policy process is re-evaluating the process through which the 'evidence' is constructed. While evidence-based policy tends to be portrayed as a neutral and objective policy tool, the very act of evidence selection is, in itself, affected by hierarchical structures governing knowledge systems. The current dominance of RCTs in evidence construction comes with the risk of making development projects blind to both stakeholder perspectives and social science theories on 'how and why' change may happen. Against this background, we argued that theories of change may be useful instruments, if applied with caution, encouraging broad and deep participation of different level stakeholders as well as offering opportunities to engage with a variety of social science disciplines throughout the research cycle.

Considering that evidence-based policy and the prevalence of RCTs were modelled on the 'gold standard' of medical experiments, perhaps it is again within the realm of natural science that social scientists should look for inspiration for the way forward. In a recent study in *the Lancet*, Jones and Podolsky (2015, p. 1503) write that

The past several years have seen increasing calls for an ecumenical approach to clinical research, with more flexible standards for what counts as acceptable study designs. Physicians have developed new methods to extract robust analyses from patient registries and from the ever-growing databases provided by electronic medical records. Will this erode the status of RCTs as a gold standard?

References

- APPAM (2015). Association for Public Policy Analysis and Management 37 Annual Research Conference: The Golden Age of Evidence-Based Policy, November 12–14, Miami, Florida, the US.
- Astbury, B. and Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), 363–381.
- Aune, J. B. (2000). Logical framework approach and PRA - mutually exclusive or complementary tools for project planning? *Development in Practice*, 10(5), 687–690.

- Baguios, A. (2019). It's time to decolonise project management in the aid sector. Aid Re-imagined for *Medium*. <https://medium.com/@aidreimagined/its-time-to-decolonise-project-management-in-the-aid-sector-da1aa30c5eee>.
- Brest, P. (2010). The power of theories of change. *Stanford Social Innovation Review*, 8(2), 47–51.
- Brown, M. (2020). Unpacking the theory of change. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/unpacking_the_theory_of_change?fbclid=IwAR2GsxxaLimnwQmG3doQJWTDra7m-rhbTJkfnWIEAQN6XHwLuP0EtUzOwLo.
- Büscher, B. (2014). Selling success: Constructing value in conservation and development. *World Development*, 57, 79–90.
- Cartwright, N. (2007). Are RCTs the Gold standard? *BioSocieties*, 2(1), 11–20.
- Clegg, S. (2007). Evidence-based practice in educational research: A critical realist critique of systematic review. *British Journal of Sociology of Education*, 26(3), 415–428.
- Coote, A., Allen, J. and Woodhead, D. (2004). Finding out what works. Building knowledge about complex, community-based initiatives. King's Fund Publications. https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/finding-out-what-works-community-based-initiatives-nov04.pdf.
- Curtis, D. and Poon, Y. (2009). Why a managerialist pursuit will not necessarily lead to achievement of MDGs. *Development in Practice*, 19(7), 837–848.
- Dale, R. (2003). The logical framework: An easy escape, a straitjacket, or a useful planning tool? *Development in Practice*, 13(1), 57–70.
- de Souza Leão, L. and Eyal, G. (2019). The rise of randomized controlled trials (RCTs) in international development in historical perspective. *Theory and Society*, 48, 383–418.
- Donovan, K. P. (2018). The rise of the randomistas: On the experimental turn in international aid. *Economy and Society*, 47(1), 27–58.
- Ferguson, J., Huysman, M. and Soekijad, M. (2010). Knowledge management in practice: Pitfalls and potentials for development. *World Development*, 38(12), 1797–1810. doi:10.1016/j.worlddev.2010.05.004
- Gasper, D. (2000). Evaluating the “logical framework approach” towards learning-oriented development evaluation. *Public Administration and Development*, 20(1), 17–28.
- Hersoug B. (1996). Logical framework analyses in an illogical world. *Forum for Development Studies*, 2, 377–404.
- Honig, D. (2019). *Navigation by judgment: Why and when top-down management of foreign aid doesn't work*. Oxford: Oxford University Press.
- Jones, D. S. and Podolsky, S. H. (2015). The art of medicine. The history and fate of the gold standard. *The Lancet*, 385, 1502–1503.
- Klerkx, L., Aarts, N. and Leeuwis, C. (2010). Adaptive management in agricultural innovation systems: The interactions between innovation networks and their environment. *Agricultural Systems*, 103, 390–400.
- Kumar, S. and Corbridge, S. (2002). Programmed to fail? Development projects and the politics of participation. *Journal of Development Studies*, 39(2), 73–103.
- Leeuwis, C. (2000) Re-conceptualizing participation for sustainable rural development. Towards a negotiation approach. *Development and Change*, 31, 931–959.
- Leigh, A. (2018). *Randomistas. How radical researchers are changing our world*. New Haven: Yale University Press.
- Leonhardt, D. (2014). The quiet movement to make the government fail less often. *The New York Times* – The Upshot 2014/07/15. <https://www.nytimes.com/2014/07/15/upshot/the-quiet-movement-to-make-government-fail-less-often.html>
- Mason, P. and Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation*, 13(2), 151–170.

- Murdach, A. D. (2010). What good is soft evidence? *Social Work*, 55(4), 309–316.
- Oakley, A. (2000). *Experiments in knowing: Gender and method in the social sciences*. Cambridge: Polity Press.
- Pahlman, K. (2014). A critical examination of the idea of evidence-based policymaking. *ANU Undergraduate Research Journal*, 6, 83–94.
- Prinsen, G. and Nijhof, S. (2015). Between logframes and theory of change: Reviewing debates and a practical experience. *Development in Practice*, 25(2), 234–246.
- Ramalingam, B., Laric, M. and Primrose, J. (2014). *From best practice to best fit*. Understanding and navigating wicked problems in international development. Overseas Development Institute. <https://www.odi.org/publications/8571-best-practice-best-fit-understanding-and-navigating-wicked-problems-international-development>.
- Stein, D. and Valters, C (2012). Understanding ‘Theory of Change’ in international development. A review of existing knowledge. Justice and Security Research Programme, London School of Economics and Political Science.
- Sullivan, H. and Stewart, M. (2006). Who owns the theory of change? *Evaluation*, 12(2), 179–199.
- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In: Connell, J. P. (eds.). *New approaches to evaluating community initiatives. concepts, methods, and contexts. Roundtable on comprehensive community initiatives for children and families*. Aspen: Aspen Institute, 65–92.